SIPTA Online School 2020, University of Liverpool Institute for Risk and Uncertainty

Game-theoretic foundations for statistical testing and imprecise probabilities

Remote lectures by **Glenn Shafer**. December 9th and 10th, 2020.

Lecture 2. Betting protocols.

Reading: Borel's law of large numbers, Chapter 1 of *Game-Theoretic Foundations for Probability and Finance*, by Glenn Shafer and Vladimir Vovk, Wiley, 2019

Slides revised January 23, 2021

1. Game theory and betting protocols
2. Testing a probability distribution
3. Testing a probability forecaster
4. One-sided offers & suboptimal bets
5. Using signals
6. Testing market efficiency

1. Game theory and betting protocols

Die Frage, ob die Anfangsposition p_0 bereits für eine der spielenden Parteien ein “Gewinnstellung” ist, steht noch offen. Mit ihrer exacten Beantwortung würde freilich das Schach den Charakter eines Spieles überhaupt verlieren.

Ernst Zermelo, 1913

To make testing-by-betting rigorous,
we must define a game.

This means specifying

1. the players,
2. how they move (simultaneously? in order?),
3. the information available to them,
4. their goals.

Economics and most other fields that use game theory use games of imperfect information. Players move simultaneously.

- von Neumann's minimax theorem
- Nash equilibrium

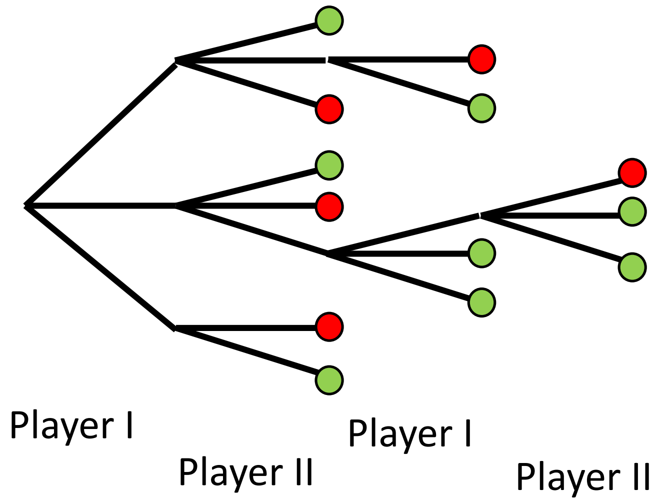
We use a *narrow but older* part of game theory:

two-player games of perfect information.

We usually ask whether one player can achieve a certain goal. The others being a team, this makes the game a two-player game.

Specify:

1. The players
2. Rules for moving
3. Information the players have when they move
4. Rule for winning (or goal for each player)



Player I wins at red.
Player II wins at green.

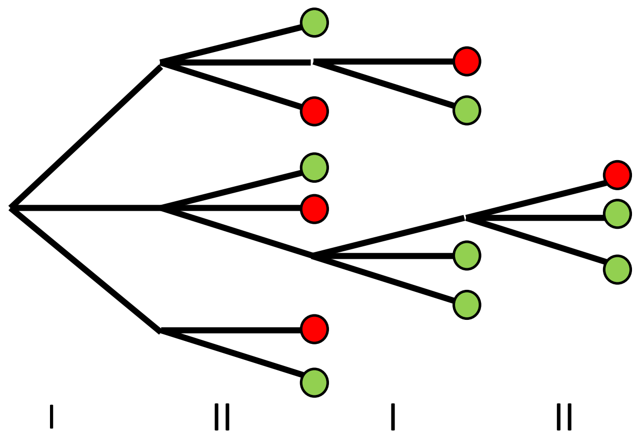
Who has a winning strategy?

Zermelo’s theorem (1913):

In a two-person perfect-information game that always has a winner, one of the players has a winning strategy.

Zermelo's theorem:

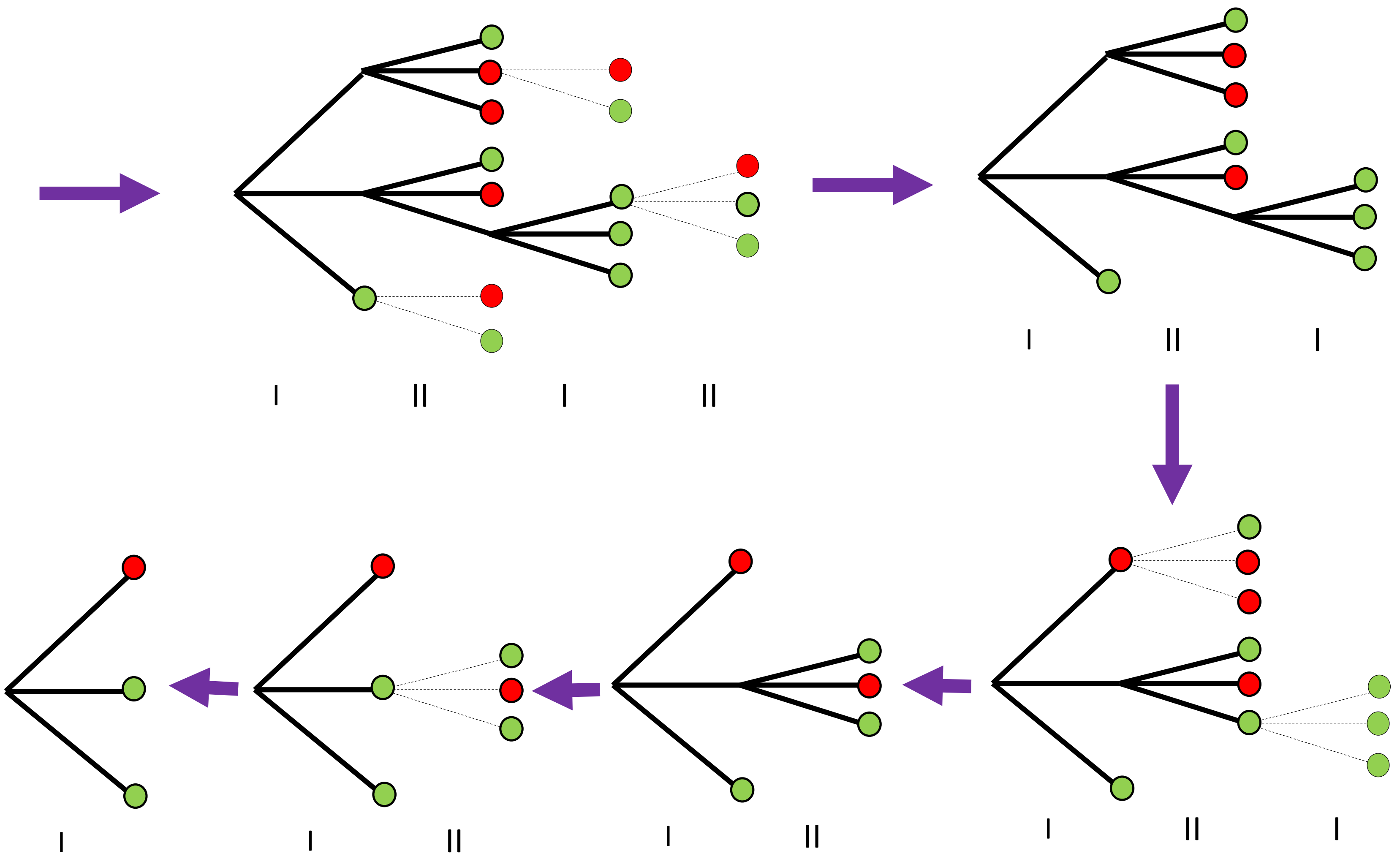
In a combinatorial game, one of the players has a winning strategy.



Player I wins at red.

Player II wins at green.

PROOF BY BACKWARD RECURSION



Ernst Zermelo
1871-1953
German logician

Zermelo's theorem (1913):

In two-person perfect-information game that always has winner, one player has a winning strategy.

Generalized in 1990 to games with infinite horizon (p. 90 of *Game-Theoretic Foundations*).

Lecture 3: Theorems that say something happens with high probability (e.g., law of large numbers) become <u>theorems in game theory</u>.

- Equate P(A) = 1-α with bettor having strategy that multiplies capital by 1/α unless A happens.
- Equate P(A) = 1 with bettor having strategy that multiplies capital infinitely unless A happens.

Zermelo's theorem:

In a two-person perfect-information game that always has a winner, one of the players has a winning strategy.

In two-person perfect information games, pure strategies are used to prove theorems.

But you can test by betting without a strategy!

Proving probability theorems is one of many ways of using testing by betting.

To make testing by betting rigorous, we must define a game.

This means specifying
1. the players
2. how they move (simultaneously? in order?),
3. the information available to them,
4. their goals.

We use betting protocols to specify 1, 2, and 3.
- The players move in order.
- They see each other's moves as they are made (perfect information).
- Each may have other information (specified by the protocol or not).

We may vary the players' goals without changing the protocol.

Many of the following examples of betting protocols are in the RSS paper studied in the first lecture.

Others are in *Game-Theoretic Foundations*.

2. Testing a probability distribution

... the odds are now only 8 to 1 against a system of deviations as improbable or more improbable than this one.

Karl Pearson, 1900

Testing a probability distribution P on $\mathcal{Y}$

Skeptic announces $S : \mathcal{Y} \to [0, \infty)$ such that $\mathbf{E}_P(S) = 1$.
Reality announces $y \in \mathcal{Y}$.
$\mathcal{K} := S(y)$.

Betting protocol

- S is the *bet*.
- $S(y)$ is the *betting score*.
- Skeptic is testing the probability distribution P.
- Skeptic is trying to make $S(y)$ large, but his real goal is to see whether P and Reality agree.

Testing a probability distribution P on $\mathcal{Y}$

Skeptic announces $S : \mathcal{Y} \to [0, \infty)$ such that $\mathbf{E}_P(S) = 1$.
Reality announces $y \in \mathcal{Y}$.
$\mathcal{K} := S(y)$.

The probability space $(P, \mathcal{Y})$ is arbitrary.

- y can be multi-dimensional.
- P can the probability distribution for a stochastic process $Y = (Y_1, \ldots, Y_N)$.

Testing a probability distribution P on $\mathcal{Y}$

Skeptic announces $S : \mathcal{Y} \to [0, \infty)$ such that $\mathbf{E}_P(S) = 1$.
Reality announces $y \in \mathcal{Y}$.
$\mathcal{K} := S(y)$.

This is a *perfect information protocol*:

- The players move sequentially.
- Each sees the other's move as it is made.

Testing a probability distribution P on $\mathcal{Y}$

Skeptic announces $S : \mathcal{Y} \to [0, \infty)$ such that $\mathbf{E}_P(S) = 1$.
Reality announces $y \in \mathcal{Y}$.
$\mathcal{K} := S(y)$.

Special case: $\mathcal{Y} = \{0, 1\}$.
Think of 1 as Heads and 0 as Tails.

Testing a probability p for $y = 1$

Skeptic announces $s_0, s_1 \in [0, \infty)$ such that $ps_1 + (1-p)s_0 = 1$.
Reality announces $y \in \{0, 1\}$.
$\mathcal{K} := ys_1 + (1 - y)s_0$.

Special case: $\mathcal{Y} = \{0,1\}$.
Think of 1 as Heads and 0 as Tails.

Testing a probability p for $y = 1$

Skeptic announces $s_0, s_1 \in [0,\infty)$ such that $ps_1 + (1-p)s_0 = 1$.
Reality announces $y \in \{0,1\}$.
$\mathcal{K} := ys_1 + (1-y)s_0$.

Another way of stating this protocol

Testing a probability p for $y = 1$

Skeptic announces M such that $\frac{-1}{1-p} \leq M \leq \frac{1}{p}$.
Reality announces $y \in \{0,1\}$.
$\mathcal{K} := 1 + M(y-p)$.

The two versions are related by

$$M = \frac{s_1 - 1}{1-p} = \frac{1 - s_0}{p}.$$

The two versions when $p = \frac{1}{2}$

Testing the probability $\frac{1}{2}$ for $y = 1$

Skeptic announces $s_0, s_1 \in [0, \infty)$ such that $(s_1 + s_0)/2 = 1$.

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

$\mathcal{K} := y s_1 + (1 - y) s_0$.

Testing the probability $\frac{1}{2}$ for $y = 1$

Skeptic announces M such that $-2 \leq M \leq 2$.

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

$\mathcal{K} := 1 + M(y - \frac{1}{2})$.

Yet another variation: Use $\{-1, 1\}$ instead of $\{0, 1\}$ as $\mathcal{Y}$.

Testing a probability $\frac{1}{2}$ for $y = 1$

Skeptic announces M such that $-1 \leq M \leq 1$.

Reality announces $y \in \{-1, 1\}$.

$\mathcal{K} := 1 + M y$.

3. Testing a probability forecaster

...the observation ... that many chaotic empirical phenomena ... exhibit stable [relative frequencies] ... arises as a consequence of our methodological practice of first removing noticeable regularities (e.g., drifts, cycles, etc.) from the data ...

Terrence L. Fine, 1976

Is P a good description of reality?

Testing a probability distribution P on $\mathcal{Y}$

Skeptic announces $S : \mathcal{Y} \to [0, \infty)$ such that $\mathbf{E}_P(S) = 1$.

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

$\mathcal{K} := S(y)$.

Is Forecaster a good forecaster?

Testing a forecaster

Forecaster announces a probability distribution P on $\mathcal{Y}$.

Skeptic announces $S : \mathcal{Y} \to [0, \infty)$ such that $\mathbf{E}_P(S) = 1$.

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

$\mathcal{K} := S(y)$.

Psychological shift:
P belongs to Forecaster,
not to Reality.

Testing a forecaster

Forecaster announces a probability distribution P on $\mathcal{Y}$.
Skeptic announces $S : \mathcal{Y} \to [0, \infty)$ such that $\mathbf{E}_P(S) = 1$.
Reality announces $y \in \mathcal{Y}$.
$\mathcal{K} := S(y)$.

Testing a forecaster over time

$\mathcal{K}_0 := 1$.
FOR $n = 1, 2, \ldots, N$:
Forecaster announces a probability distribution P_n on $\mathcal{Y}$.
Skeptic announces $S_n : \mathcal{Y} \to [0, \infty)$ such that $\mathbf{E}_{P_n}(S_n) = \mathcal{K}_{n-1}$.
Reality announces $y_n \in \mathcal{Y}$.
$\mathcal{K}_n := S_n(y_n)$.

This is outside standard probability.
We have no joint distribution for $y_1, \ldots y_N$.

Testing a forecaster over time

$\mathcal{K}_0 := 1$.
FOR $n = 1, 2, \ldots, N$:
  Forecaster announces a probability distribution P_n on $\mathcal{Y}$.
  Skeptic announces $S_n : \mathcal{Y} \to [0, \infty)$ such that $\mathbf{E}_{P_n}(S_n) = \mathcal{K}_{n-1}$.
  Reality announces $y_n \in \mathcal{Y}$.
  $\mathcal{K}_n := S_n(y_n)$.

Compare to testing a probability distribution P for a stochastic process $Y = (Y_1, \ldots, Y_n)$.

Probability distribution for stochastic process is a kind of strategy for Forecaster.

Probability distributions don't tell Reality what to do. They tell Forecaster what to do.

Testing a stochastic process

$\mathcal{K}_0 := 1$.
FOR $n = 1, 2, \ldots, N$:
  Skeptic announces $S_n : \mathcal{Y} \to [0, \infty)$ such that $\mathbf{E}_P(S_n(Y_n)|y_1, \ldots, y_{n-1}) = \mathcal{K}_{n-1}$.
  Reality announces $y_n \in \mathcal{Y}$.
  $\mathcal{K}_n := S_n(y_n)$.

4. One-sided offers & suboptimal bets

> … it is always possible for the individual to choose whether to bet at the extreme rates, or only arbitrarily closely, in such a way that he never accepts a bet which he can only lose.
>
> Peter M. Williams, 1975

One-sided betting offer

Skeptic is only allowed to bet on $y = 1$.
(Forecaster: $y = 0$ more likely than not.)

Testing the probability $\frac{1}{2}$ for $y = 1$

Skeptic announces M such that $0 \leq M \leq 2$.
Reality announces $y \in \{0, 1\}$.
$\mathcal{K} := 1 + M(y - \frac{1}{2})$.

Bets suboptimal for Skeptic

Skeptic can give up money.
(Calculation too difficult?)

Testing a forecaster over time

$\mathcal{K}_0 := 1$.
FOR $n = 1, 2, \ldots, N$:
Forecaster announces a probability distribution P_n on $\mathcal{Y}$.
Skeptic announces $S_n : \mathcal{Y} \to [0, \infty)$ such that $\mathbf{E}_{P_n}(S_n) \leq \mathcal{K}_{n-1}$.
Reality announces $y_n \in \mathcal{Y}$.
$\mathcal{K}_n := S_n(y_n)$.

5. Using signals

> The idea of regression is usually introduced in connection with the theory of correlation, but it is in reality a more general, and, in some respects a simpler idea, and the regression coefficients are of interest and scientific importance in many classes of data where the correlation coefficient, if used at all, is an artificial concept of no real utility.
>
> R. A. Fisher, 1925

Introducing auxiliary information x, sometimes known as

- *signal* in engineering,
- *independent variable* in statistics,
- *object* in machine learning.

Testing a forecaster over time

$\mathcal{K}_0 := 1$.
FOR $n = 1, 2, \ldots, N$:
  Reality announces $x_n \in \mathcal{X}$.
  Forecaster announces a probability distribution P_n on $\mathcal{Y}$.
  Skeptic announces $S_n : \mathcal{Y} \to [0, \infty)$ such that $\mathbf{E}_{P_n}(S_n) = \mathcal{K}_{n-1}$.
  Reality announces $y_n \in \mathcal{Y}$.
  $\mathcal{K}_n := S_n(y_n)$.

Testing a forecaster over time

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

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

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

  Forecaster announces a probability distribution P_n on $\mathcal{Y}$.

  Skeptic announces $S_n : \mathcal{Y} \to [0, \infty)$ such that $\mathbf{E}_{P_n}(S_n) = \mathcal{K}_{n-1}$.

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

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

What does a strategy for Forecaster in this protocol look like?

One type of strategy for Forecaster specifies for all n and all possible $x_1 y_1 \ldots x_{n-1} y_{n-1} x_n$ a probability distribution $P_{x_1 y_1 \ldots x_{n-1} y_{n-1} x_n}$.

Testing a forecaster's strategy that uses a signal

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

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

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

  Skeptic announces $S_n : \mathcal{Y} \to [0, \infty)$ such that

$$\mathbf{E}_{P_{x_1 y_1 \ldots x_{n-1} y_{n-1} x_n}}(S_n) = \mathcal{K}_{n-1}.$$

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

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

Again we seem to be outside standard probability theory.

Psychological divergence between probabilists and mathematical statisticians

By late 20th century, mathematical probability meant

- Kolmogorov (axioms, abstract Lebesgue integration, conditional expectation) plus
- Doob (add filtration).

In the resulting "folk picture",

- all probabilities live within a single filtered probability space $(\Omega, \mathbf{P}, (\mathcal{F}_t)_{t\in\mathbb{R}})$,
- Ω is so huge that the complete actual evolution of the universe is described by a single $\omega \in \Omega$.

The probabilists' folk picture has infiltrated some applied work, <u>but</u>

- applied statistics often retains the picture where probabilities merely describe or forecast small worlds, as in Fisher's fixed-x regression,
- statisticians continue to combine p-values à la Fisher even though they do not believe in a global probability distribution for what experiments would have been conducted,
- those statisticians who consider probability subjective tend to reject the global nature of the folk picture,
- workers in "imprecise probability" now pursue notions of subjective probability that use less than a probability measure,
- game-theoretic probability also moves away from complete probability measures.

6. Testing market efficiency

,,,market efficiency per se is not testable. It must be tested jointly with some model of equilibrium, an asset-pricing model.

Eugene Fama, 1991

Remember our protocol for the probability $\frac{1}{2}$ for Heads when Heads is coded as 1 and Tails as -1:

Testing a probability $\frac{1}{2}$ for $y = 1$
Skeptic announces M such that $|M| \leq 1$.
Reality announces $y \in \{-1, 1\}$.
$\mathcal{K} := 1 + My$.

Generalize this from one to many flips:

$\mathcal{K}_0 := 1$.
FOR $n = 1, 2, \ldots, N$:
Skeptic announces M_n such that $|M_n| \leq \mathcal{K}_{n-1}$.
Reality announces $y_n \in \{-1, 1\}$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n y_n$.

$\mathcal{K}_0 := 1.$
FOR $n = 1, 2, \ldots, N$:
Skeptic announces M_n such that $|M_n| \le \mathcal{K}_{n-1}$.
Reality announces $y_n \in \{-1, 1\}$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n y_n$.

Now make one small further change: change Reality's move space from $\{-1, 1\}$ to $[-1, 1]$.

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

Here Skeptic is testing a forecaster who gives the estimate 0 to successive quantities $y_1, \ldots, y_n$ that are bounded between -1 and 1.

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

Let's make it more interesting and allow Forecaster to vary his estimate:

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

Protocols of this type can be used to study problems often called "non-parametric".

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

With one more twist, we get a protocol for testing the efficiency of a financial market. Today's price for a financial security is an estimate of tomorrow's:

$\mathcal{K}_0 := 1$.
Market announces $y_0 \in (0, \infty)$.
FOR $n = 1, 2, \ldots, N$:
  Skeptic announces M_n such that $|M_n| \le \mathcal{K}_{n-1}$.
  Market announces $y_n \in (\max(0, y_{n-1} - 1), y_{n-1} + 1]$.
  $\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(y_n - y_{n-1})$.

WP 1, 3, 44, 47 at `www.probabilityandfinance.com`.

- A stock price cannot be negative.
- We are assuming a limit on its change in one day.