Themes for a textbook

Glenn Shafer

CWI E-Day

November 8, 2024

I revised these slides on November 11 in response to some of the questions and issues raised in the discussion. See in particular slides 17 and 24.

An introduction to game-theoretic statistics

Began book in late 2022, inspired by Eindhoven meeting.

Taught from earlier drafts, Spring 2023 and Spring 2024.

Want your feedback on the most novel themes.

Ten Themes

1. Forecaster, Skeptic, and Reality easier to teach than Type I and Type II error
2. Keep it simple: no measure theory, finite horizon
3. Distinguish Forecaster's *forecasts* from Statistician's *conclusions*
4. Cournot's principle. How it became paradoxical. Resolving the paradox.
5. Arguments for Kelly testing
6. The subjective and the objective: Forecaster, Skeptic, Statistician
7. Compare the *betting-score* and *p-value* scales.
8. Replace statistical models with Oracle.
9. What is important in the generalization from standard probability?
10. Teach limits of statistics (and perils of gambling) while teaching statistics.

Table of Contents

1. Forecaster, Skeptic, and Reality easier than Type I and Type II error
2. Keep it simple: No measure theory, finite horizon

Protocol 6.1: General probability forecasting

:bf⟩ PLAYERS: Skeptic, Forecaster, Reality

PARAMETERS: $N \in \mathbb{N}$, probability space $\mathcal{Y}$

Skeptic announces $s_0 \in \mathbb{R}$.

For $n = 1, \ldots, N$:

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

Skeptic announces variable G_n on $\mathcal{Y}$ with finite $\mathbf{E}_{P_n}(G_n)$.

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

$s_n := s_{n-1} + G_n(y_n) - \mathbf{E}_{P_n}(G_n)$.

Protocol 6.1: General probability forecasting

:bf⟩ PLAYERS: Skeptic, Forecaster, Reality

PARAMETERS: $N \in \mathbb{N}$, probability space $\mathcal{Y}$

Skeptic announces $s_0 \in \mathbb{R}$.

For $n = 1, \dots, N$:

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

Skeptic announces variable G_n on $\mathcal{Y}$ with finite $\mathbf{E}_{P_n}(G_n)$.

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

$s_n := s_{n-1} + G_n(y_n) - \mathbf{E}_{P_n}(G_n)$.

Betting with play money.

Protocol is imagined by Statistician.

Protocol 6.1: General probability forecasting

:bf⟩ PLAYERS: Skeptic, Forecaster, Reality
PARAMETERS: $N \in \mathbb{N}$, probability space $\mathcal{Y}$
Skeptic announces $s_0 \in \mathbb{R}$.
For $n = 1, \ldots, N$:
Forecaster announces probability distribution P_n on $\mathcal{Y}$.
Skeptic announces variable G_n on $\mathcal{Y}$ with finite $\mathbf{E}_{P_n}(G_n)$.
Reality announces $y_n \in \mathcal{Y}$.
$s_n := s_{n-1} + G_n(y_n) - \mathbf{E}_{P_n}(G_n)$.

Protocol is imagined by Statistician.

Bet with play money.

Statistician tells Skeptic what to say.
Sometimes she also tells Forecaster what to say.

Sometimes she assigns Skeptic or Forecaster strategies at the outset.

Protocol 6.1: General probability forecasting

:bf⟩ PLAYERS: Skeptic, Forecaster, Reality
PARAMETERS: $N \in \mathbb{N}$, probability space $\mathcal{Y}$
Skeptic announces $s_0 \in \mathbb{R}$.
For $n = 1, \ldots, N$:
Forecaster announces probability distribution P_n on $\mathcal{Y}$.
Skeptic announces variable G_n on $\mathcal{Y}$ with finite $\mathbf{E}_{P_n}(G_n)$.
Reality announces $y_n \in \mathcal{Y}$.
$s_n := s_{n-1} + G_n(y_n) - \mathbf{E}_{P_n}(G_n)$.

Statistician	she
Forecaster	he
Skeptic	he
Reality	she

Protocol 6.1: General probability forecasting

:bf⟩ PLAYERS: Skeptic, Forecaster, Reality
PARAMETERS: $N \in \mathbb{N}$, probability space $\mathcal{Y}$
Skeptic announces $s_0 \in \mathbb{R}$.
For $n = 1, \ldots, N$:
Forecaster announces probability distribution P_n on $\mathcal{Y}$.
Skeptic announces variable G_n on $\mathcal{Y}$ with finite $\mathbf{E}_{P_n}(G_n)$.
Reality announces $y_n \in \mathcal{Y}$.
$s_n := s_{n-1} + G_n(y_n) - \mathbf{E}_{P_n}(G_n)$.

Sometimes I change Reality's name to *Informant*.

Either way, this player announces facts that Statistician does not control.

Protocol 6.1: General probability forecasting

:bf⟩ PLAYERS: Skeptic, Forecaster, Reality
PARAMETERS: $N \in \mathbb{N}$, probability space $\mathcal{Y}$
Skeptic announces $s_0 \in \mathbb{R}$.
For $n = 1, \ldots, N$:
Forecaster announces probability distribution P_n on $\mathcal{Y}$.
Skeptic announces variable G_n on $\mathcal{Y}$ with finite $\mathbf{E}_{P_n}(G_n)$.
Reality announces $y_n \in \mathcal{Y}$.
$s_n := s_{n-1} + G_n(y_n) - \mathbf{E}_{P_n}(G_n)$.

Fundamental principle for testing by betting. Successive bets against a forecaster that begin with unit capital and never risk more discredit the forecaster to the extent that the bettor's cumulative capital is large.

3. Distinguish Forecaster's *forecasts* from Statistician's *conclusions*

Fundamental principle for game-theoretic induction. Statistician may conclude that a forecaster who has consistently withstood certain test strategies in the past will withstand similar test strategies in the future.

Fundamental principle for game-theoretic induction. Statistician may conclude that a forecaster who has consistently withstood certain test strategies in the past will withstand similar test strategies in the future.

A test strategy is a strategy for Skeptic that

- begins with unit capital and
- never makes a move that risks making its capital negative.

4. Cournot's principle

How it became paradoxical.

Resolving the paradox.

Cournot: Events of high probability are practically certain.

Principle considered fundamental by Aquinas, Bernoulli, Condorcet, Borel, Levy, Kolmogorov, Ville, Doob, etc.

Critics evoke the lottery paradox: A small probability event always happens.

Why was the lottery paradox overlooked before the 1960s?

1. Earlier authors thought about "certainty" differently.
2. Earlier authors did not begin with a probability measure.

To modernize Cournot's principle,

1. **replace certainty with conclusion (or prediction)** and
2. **use only simple high-probability forecasts as predictions**.

This works best with game-theoretic probability.

Why was the lottery paradox overlooked before the 1960s?

1. Earlier authors thought about "certainty" differently.
2. Earlier authors did not begin with a probability measure.

To modernize Cournot's principle,

1. **replace certainty with prediction** and
2. **use only simple high-probability forecasts as predictions**.

This works best with game-theoretic probability.

Fundamental principle for game-theoretic induction. Statistician may conclude that a forecaster who has consistently withstood certain test strategies in the past will withstand similar test strategies in the future.

Wouter asked during my talk, what is predicted?

Although the events we "predict" in everyday speech are future events, the global events that Statistician predicts are not necessarily future events; it is the predictions that lie in the future. These predictions will be based on forecasts that Forecaster makes in the future, or least that Statistician notices in the future.

Statistician may never observe whether her predictions are correct, and in some cases the events predicted are not of direct interest. What is interesting is other facts that can be inferred from the assumption that they are correct, together with other arguments. This is statistical inference, which we will study in Chapter 10.

Statistician's predictions can also be used in decision making. This is possible when we can combine Forecaster's forecasts with utilities in a way that Statistician can calculate predicted average utilities for a series of future decisions.

5. Arguments for Kelly testing

Kelly's argument:

Taking logs makes the process additive.
So the law of large numbers applies.

Another argument when testing by betting:

Human perception is logarithmic.

The difference between 5 and 25 is huge.
The difference between 105 and 125 is neglible.

6. The subjective and the objective: Forecaster, Skeptic, Statistician

Forecaster wants to resist Skeptic's long-run tests.

Resisting Skeptic's long-run tests includes matching Reality's frequencies and averages.

Skeptic may use subjective probabilities to make Kelly tests.

Statistician might believe that Forecaster's forecasts will succeed (prove to be objective).

Statistician might have subjective probabilities matching Skeptic's (or Forecaster's).

7. Compare the *betting-score* and *p-value* scales

My recommendation for interpreting betting scores:

≥ 3.5	mild discredit
≥ 10	strong discredit

Shrink inverse of a p-value to make it a betting score:

$$p \mapsto \sqrt{1/p} - 1.$$

Table 4.1: Some p-values commonly used as thresholds and corresponding betting scores via the default mapping $p \mapsto \sqrt{1/p} - 1$.

p-value p	0.1	0.05	0.01	0.005	0.001	0.0001	0.00001
$\sqrt{1/p} - 1$	2.2	3.5	9	13	31	99	315

Rule of thumb: $p \mapsto \sqrt{1/p} - 1$ $s \mapsto (1+s)^{-2}$

Forecaster announces normal distribution with mean μ.
Skeptic tests with normal distribution with mean υ and the same variance.

s := betting score

s^* := implied target

$$\lambda := \frac{|\nu - \mu|}{\sigma}$$

λ	s^*	s = 2.5	3.5	7	10	30	50
1.25	2.2	0.087	0.052	0.015	0.0068	0.00041	0.000087
1.5	3.1	0.087	0.056	0.020	0.011	0.0013	0.00039
2	7.4	0.072	0.052	0.024	0.016	0.0035	0.0016
2.5	23	0.053	0.040	0.021	0.015	0.0045	0.0024
3	90	0.036	0.028	0.016	0.012	0.0042	0.0025
4	2981	0.013	0.010	0.0064	0.005	0.0022	0.0015
$(1+s)^{-2}$		0.082	0.049	0.016	0.0083	0.0010	0.00038

The rule of thumb is pretty good when Forecaster and Skeptic differ by a few σ.

Table 4.5: Three mappings from p-values to betting scores, applied to some conventional thresholds for p-values. Here $\kappa_0 = -1/\ln(0.05)$; this is the value of κ that maximizes $\kappa(0.05)^{\kappa-1}$.

p	$\frac{1}{\sqrt{p}} - 1$	$\kappa_0 p^{\kappa_0 - 1}$	$\frac{1-p-p\ln p}{p(-\ln p)^2}$
0.05	3.5	2.5	1.8
0.01	9	7.2	4.5
0.005	13	11	6.9
		Volodya	Aaditya

Volodya and Aaditya sought to make the betting score as close to *1/p* as possible for very small *p*.

$$\int_0^1 \kappa p^{\kappa-1} d\kappa = \frac{1-p-p\ln p}{p(-\ln p)^2}$$

Peter remarked, during my talk, that no general mapping from p-values to betting scores holds across problems.

I agree. I also agree with Harold Jeffreys that if you get a Bayes factor (or betting score) larger than 100, it does not matter (so far as the intuitive message about practical certainty is concerned) how much larger it is.

We should use a rule of thumb for converting a p-value into a betting score only when we are given a p-value and do not have the time or information needed to calculate a betting score.

8. Replace statistical models with Oracle.

A game-theoretic model

Write N_{μ,σ^2} for the normal distribution with mean μ and variance σ^2.

Protocol 6.24: Regression model

PLAYERS: Skeptic, Informant, Reality
PARAMETERS: $N \in \mathbb{N}$, $(\beta_0, \beta_1, \ldots, \beta_p) \in \mathbb{R}^{p+1}$, and $\sigma^2 \in (0, \infty)$
 Skeptic announces $s_0 \in \mathbb{R}$.
 Informant announces $(x_1, \ldots, x_p) \in \mathbb{R}^p$.
 Skeptic announces variable G on $\mathbb{R}$.
 Reality announces $y \in \mathbb{R}$.
 $s := s_0 + G(y) - \mathbf{E}_{N_{\beta_0+\beta_1 x_1+\cdots+\beta_p x_p,\sigma^2}}(G)$.

Statistician imagines a game where she does not see all the moves.

Protocol 6.25: Regression with Oracle

PLAYERS: Skeptic, Informant, Reality
PARAMETERS: $N \in \mathbb{N}$
 Oracle announces $(\beta_0, \beta_1, \ldots, \beta_p) \in \mathbb{R}^{p+1}$, and $\sigma^2 \in (0, \infty)$
 Skeptic announces $s_0 \in \mathbb{R}$.
 Informant announces $(x_1, \ldots, x_p) \in \mathbb{R}^p$.
 Skeptic announces variable G on $\mathbb{R}$.
 Reality announces $y \in \mathbb{R}$.
 $s := s_0 + G(y) - \mathbf{E}_{N_{\beta_0+\beta_1 x_1+\cdots+\beta_p x_p,\sigma^2}}(G)$.

What statistician sees

9. Which part of game-theoretic probability's generalization of standard probability is most important?

1. Forecasts need not be complete probability distributions.
2. Game-theoretic law of large numbers applies even when Forecaster is a free agent.
3. Independent variables can be determined in the course of an experiment (or series of experiments) without having probabilities.

1. Forecasts need not be complete probability distributions.

Protocol 6.17: Mean-variance forecasting

mv⟩ PLAYERS: Skeptic, Forecaster, Reality

Skeptic announces $s_0 \in \mathbb{R}$.

Forecaster announces $\mu \in \mathbb{R}$ and $\sigma^2 \in [0, \infty)$.

Skeptic announces $g \in \mathbb{R}$ and $h \in [0, \infty)$.

Reality announces $y \in \mathbb{R}$.

$s := s_0 + g(y - \mu) + h((y - \mu)^2 - \sigma^2)$.

2. Game-theoretic law of large numbers applies even when Forecaster is a free agent.

Protocol 9.3: Binary probability forecasting (Protocol 6.8) again

re⟩ PLAYERS: Skeptic, Forecaster, Reality
PARAMETER: $N \in \mathbb{N}$
Skeptic announces $s_0 \in \mathbb{R}$.
For $n = 1, \ldots, N$:
Forecaster announces $p_n \in [0,1]$.
Skeptic announces $g_n \in \mathbb{R}$.
Reality announces $y_n \in \{0,1\}$.
$s_n := s_{n-1} + g_n(y_n - p_n)$.

Proposition 9.3

h⟩ For any $\epsilon > 0$ and any value of the parameter N in Protocol 9.3,

$$\overline{\mathbb{P}}\left(|\bar{y}_N - \bar{p}_N| \geq \epsilon\right) \leq \frac{1}{\epsilon^2 N}$$

and

$$\overline{\mathbb{P}}\left(|\bar{y}_N - \bar{p}_N| \geq \epsilon\right) \leq 2\exp\left(-\frac{\epsilon^2 N}{4}\right). \qquad (9.5)$$ eq:

3. Independent variables can be determined in the course of an experiment (or series of experiments) without having probabilities.

Protocol 6.18: Auxiliary information

ux⟩ PLAYERS: Skeptic, Informant, Reality, Forecaster

PARAMETERS: Set $\mathcal{X}$, probability space $\mathcal{Y}$

> Skeptic announces $s_0 \in \mathbb{R}$.
> Informant announces $x \in \mathcal{X}$.
> Forecaster announces probability distribution P on $\mathcal{Y}$.
> Skeptic announces variable G on $\mathcal{Y}$ with finite $\mathbf{E}_P(G)$.
> Reality announces $y \in \mathcal{Y}$.
> $s := s_0 + G(y) - \mathbf{E}_P(G)$.

Protocol 6.26: Regression with Oracle

PLAYERS: Skeptic, Informant, Reality

PARAMETERS: $N \in \mathbb{N}$

Oracle announces $(\beta_0, \beta_1, \ldots, \beta_p) \in \mathbb{R}^{p+1}$, and $\sigma^2 \in (0, \infty)$

Skeptic announces $s_0 \in \mathbb{R}$.

Informant announces $(x_1, \ldots, x_p) \in \mathbb{R}^p$.

Skeptic announces variable G on $\mathbb{R}$ with finite $\mathbf{E}_{N_{\beta_0+\beta_1 x_1+\cdots+\beta_p x_p, \sigma^2}}(G)$.

Reality announces $y \in \mathbb{R}$.

$s := s_0 + G(y) - \mathbf{E}_{N_{\beta_0+\beta_1 x_1+\cdots+\beta_p x_p, \sigma^2}}(G)$.

Strategy for Forecaster is built into protocol.

The game-theoretic law of large numbers, applied to many rounds of this protocol, gives a high lower probability to the average y approximating the average $y - \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$.

10. Teach limits of statistics (& perils of gambling) while teaching statistics

The law of large numbers permits wide deviations.

There are many other pitfalls in prediction by induction.

Delusions of martingaling.

Ten Themes

1. Forecaster, Skeptic, and Reality easier to teach than Type I and Type II error
2. Keep it simple: no measure theory, finite horizon
3. Distinguish Forecaster's *forecasts* from Statisticians *conclusions*
4. Cournot's principle. How it became paradoxical. Resolving the paradox.
5. Arguments for Kelly testing
6. The subjective and the objective: Forecaster, Skeptic, Statistician
7. Compare the *betting-score* and *p-value* scales.
8. Replace statistical models with Oracle.
9. What is important in the generalization from standard probability?
10. Teach limits of statistics (and perils of gambling) while teaching statistics.