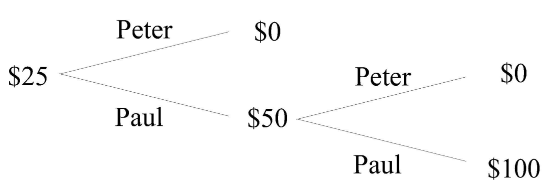
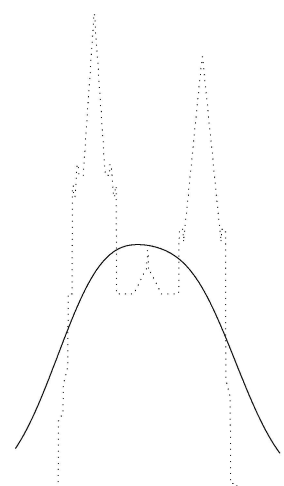
Statistical testing with optional continuation

Glenn Shafer, Rutgers University





The Game-Theoretic Probability and Finance Project

Working Paper #63

First posted August 4, 2022. Last revised June 14, 2023.

Project web site:
`http://www.probabilityandfinance.com`

Abstract

When testing a statistical hypothesis, is it legitimate to deliberate on the basis of initial data about whether and how to collect further data? The *fundamental principle for testing by betting* says yes, provided that you are testing by betting and do not risk more capital than initially committed. Standard statistical theory uses *Cournot's principle*, which does not allow such *optional continuation*. Cournot's principle can be extended to allow optional continuation when testing is carried out by multiplying likelihood ratios, but the extension lacks the simplicity and generality of testing by betting.

Keywords: optional continuation; optional stopping; Cournot's principle; game-theoretic probability; fundamental principle of testing by betting; Ville's inequality; game-theoretic probability

Contents

1 Introduction

Statistical testing requires a mathematical theory of probability together with a principle that specifies how probabilities can be discredited by observations.

- The principle used to make traditional probability theory into a theory of statistical testing is sometimes called *Cournot's principle.*[1] This principle authorizes a statistician to select an event to which a probability distribution assigns small probability and to regard its happening as evidence against the distribution.

- Game-theoretic probability is a mathematical theory in which probability distributions function as betting offers (Shafer and Vovk, 2001, 2019). To make it into a theory of statistical testing, we can use a principle that I have called *the fundamental principle for testing by betting.*[2] This principle, which is related to but distinct from Cournot's principle, authorizes a statistician to interpret success in betting against a probability distribution as evidence against the distribution.

Which of these two principles is best adapted to optional continuation? Do they both have precise formulations that assert the validity of certain statistical tests under optional continuation?

Optional continuation, as the term is used here, refers to the practice of deliberating, after seeing some initial data, whether and how to continue collecting data. Such continuation may involve observations or experiments not contemplated at the outset. It is distinguished from optional stopping, which refers only to the possibility of deciding at the outset to curtail a fully planned experiment or other study.

In this article, I conclude that the fundamental principle for testing by betting does assert the validity of optional continuation for the type of testing it considers. Cournot's principle, in its classical formulation, does not. It can be extended to assert the validity of optional continuation when testing is carried out by multiplying likelihood ratios, but the extension lacks the simplicity and generality of the fundamental principle for testing by betting.

2 Optional continuation in statistical practice and theory

Optional continuation has long been part of statistical practice. It is implicit, for example, in the idea of meta-analysis. But it has proven difficult to bring it under the purview of statistical theory.

[1] See Shafer and Vovk (2006), Shafer (2007, 2022), and the many references therein for the history of Cournot's principle. The principle has sometimes been ridiculed by philosophers, see for example Diaconis and Skyrms (2018). But it has been articulated in one way or another by a panoply of mathematicians and statisticians, including Jacob Bernoulli, Antoine-Augustin Cournot himself, Émile Borel, Andrei Kolmgorov, Richard von Mises, and Charles Stein.

[2] Vovk and I have used various names for this principle. In 2001, we called it *the fundamental interpretative hypothesis of probability* (Shafer and Vovk, 2001, pp. 5, 14, 62). In 2019, we called it *the game-theoretic version of Cournot's principle* (Shafer and Vovk, 2019, pp. 226–227).

The term "optional continuation" with the meaning used here first appeared in print in Allard Hendriksen's master's thesis at the University of Leiden (Hendriksen, 2017), written under the supervision of Peter Grünwald. Hendriksen wrote on page 3 of the thesis,

> "Optional continuation" is the practice of combining evidence of studies that were done because of promising results of previous research on the same subject.

The term has subsequently been used in other work by Günwald's machine-learning research group at CWI in Amsterdam (Grünwald et al., 2021, 2023). But as of June 13, 2023, it had not yet appeared in any of the 34 statistics journals in JSTOR.

The older term "optional stopping" was introduced by the Duke mathematician Joseph Albert Greenwood (Greenwood, 1938). Greenwood sought empirical adjustments to account for the way Joseph Rhine's laboratory was conducting and analyzing its experiments on extra-sensory perception. Rhine stopped experimenting with each subject when a success rate thought to be statistically significant was achieved, then combined the z-scores achieved by successive subjects.

Greenwood's problem was brought to wider attention in mathematical statistics by William Feller in a critique of the ESP work (Feller, 1940, pp. 272, 286–292) and in the first edition of his textbook on probability (Feller, 1950, pp. 140, 190, 197). In subsequent work in probability, "optional stopping" has referred to stopping rules that can be adopted in advance without annulling a desired property of a stochastic process, usually the property of being a martingale; see Doob (1953).

In his book on sequential analysis (Wald, 1947), Abraham Wald considered only "sequential sampling plans" chosen in advance. While allowing early stopping when there was enough evidence to make a decision, these plans specified whether or not to stop and how to continue if stopping was not mandated, all as a function of outcomes so far. In a review of the book (Barnard, 1947), George Barnard wrote that sequential analysis marked "the entry of statistical considerations into the very process of experimentation itself." We know that the process of experimentation often involves not only plans adopted in advance but also opportunistic changes in plans, based on new insights and unexpected information.

Barnard seems not to have followed up on his insight concerning the role of statistics in the process of experimentation; he does not discuss it, for example in his major article on statistical inference (Barnard, 1949). But in a subsequent article entitled "Sequential experimentation", R. A. Fisher wrote about the need for sequential deliberation in these terms (Fisher, 1952, p. 183):

> The present use of the term sequential is intended to be of a broader import than the formal use of the word as associated with the systematic procedure known as sequential analysis. The experimenter does not regard his material as wholly passive but instead looks to

> what may be learnt from it with a view to the improvement and extension of the enquiry. This willingness to learn from it how to proceed is the essential quality of sequential procedures. Wald introduced the sequential test, but the sequential idea is much older. For example, what is the policy of a research unit? It is that in time we may learn to do better and follow up our more promising results. The essence of sequential experimentation is a series of experiments each of which depends on what has gone before. For example, in a sample survey scheme, as explained by Yates, a pilot survey is intended to supply a basis for efficiently planning the subsequent stages of a survey. ...

Until the recent work on optional continuation, this insight about statistical practice has remained outside the ambit of statistical theory.

3 A betting game with optional continuation

The simplest game used in game-theoretic probability has three players: Forecaster makes probability predictions, Skeptic bets against them, and Reality announces the outcomes. In our 2001 and 2019 books, Shafer and Vovk (2001, 2019), Vovk and I discussed the role this game can play in statistics but emphasized its mathematics, proving theorems about what each player can accomplish with various strategies.

The game is a perfect-information game, in the sense that Forecaster and Skeptic move in turn and see each other's moves. We can vary the rules of the game, but we need not impose any further condition on what information any player might have or acquire in the course of the game, or how the players might collaborate. Forecaster and Skeptic might be the same person. Forecaster and Reality might be the same person.

If Forecaster keeps forecasting, Skeptic can keep betting. Forecaster need not follow a plan or strategy about what to forecast next or how to forecast it.[3] Even if Forecaster follows a strategy, Skeptic need not have a plan or strategy for when or how to bet on the forecasts. Thus optional continuation is built into the game, for both Forecaster and Skeptic. Skeptic can decide whether and how to continue selecting from Forecaster's betting offers, but Forecaster can decide what experiments or observations to make and what forecasts (perhaps probabilities) to give for them.

In our 2001 and 2019 books, Vovk and I used the example of quantum mechanics to illustrate game-theoretic probability's capacity for optional continuation; see (Shafer and Vovk, 2001, pp. 189–191) and (Shafer and Vovk, 2019, pp. 215–217). In this example, we split Forecaster into two players, Observer and Quantum Mechanics. Observer selects the experiment, and Quantum Mechanics makes the probability forecast. Formally, the game continues indefinitely, but both Observer and Skeptic can effectively stop it by making null moves.

[3] To see how probability's limit theorems can be generalized to accommodate Forecaster's freedom, see (Shafer and Vovk, 2019, §7.5).

Although optional continuation is built into the game, we need this principle to use the game in statistical testing:

Principle 1 (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 final capital is large.*[4]

In one sense, this says it all. But some elaboration may be useful:

1. The principle is *fundamental*, not the consequence of some more extensive philosophy or methodology. We do not begin by saying that the forecaster's probabilities are or should be objective, subjective, personal, "frequentist", or whatever. We are testing the forecaster qua forecaster, and so we are testing his forecasts qua forecasts; the question is only whether they are good forecasts, relative to the knowledge and skill of whoever is doing the testing.

2. The forecaster may give a probability for a single event A, a probability distribution for an outcome X, or something less than a probability or a probability distribution:

   - If the forecaster gives a probability, you may bet on either side at the corresponding odds.
   - If the forecaster gives a probability distribution for X, you may buy or sell any payoff $S(X)$ for its expected value.
   - If the forecaster gives only an estimate E of X, you may buy or sell X for E.
   - If the forecaster repeatedly gives a new probability for A or new estimate for X, say daily, you may buy or sell tomorrow's price for today's price.
   - If the forecaster gives upper and lower previsions, you may buy at the upper or sell at the lower.

3. You *begin with unit capital* only for mathematical convenience. The discredit is measured by the ratio (final capital)/(initial capital).

4. If you make several bets against the same forecaster (or the same theory or closely related theories), each starting with its own capital, then you are not allowed to report only the cases where you discredited the forecaster. Instead, you must report the overall result, the sum of your final capital over all the bets divided by the sum of your initial capital over all the bets.

5. When betting against successive forecasts, each bet uses only the capital remaining from the previous bet. You may not borrow or otherwise raise more capital in order to continue betting. This is what *never risk more* than the initial capital means.

[4] I first formulated the principle in this way in my SIPTA lectures (Shafer, 2020).

6. When you stop, you must compare your initial capital with your *final* capital. You cannot claim to have discredited the forecaster because you had reached a higher level of capital in the interim. This acknowledges the fact that you do not have the money if you keep betting and lose it.[5]

I have stated the fundamental principle for testing by betting in 26 words, then taken a page to explain it. Is the principle simple? In any case, it is coherent and teachable. In contexts where the forecasts are only single probabilities or estimates, the principle can be taught even to those who have never studied mathematical probability. Moreover, the principle builds on ideas about betting that most people acquire before ever studying mathematical probability. Too many predictions contradicted by experience discredit the person making them. If you lose too much money betting on something, you are not much of an expert about it. Etc.

4 Cournot's principle in classical form

What principles must we add to traditional probability theory to allow optional continuation?

Before answering this question, we answer a more basic question: How are we authorized to discredit a probability distribution P using observations? The classical answer is Cournot's principle: we select an event E that has small probability $P(E)$ (call E our *test event*). The probability distribution P is discredited if E happens; we prefer to believe that the probabilities are incorrect rather than think that this improbable event happened.

Principle 2 (Cournot's principle)**.** *If we specify an event E in advance, and E happens, then we may take α, the probability of E, as a measure of evidence against P. The magnitude of discredit is measured by how small α and thus how large $1/\alpha$ is.*

We may call $1/\alpha$ our *test score*:

$$\text{test score} = \begin{cases} 1/\alpha & \text{if } E \text{ happen} \\ 0 & \text{if } E \text{ does not happen.} \end{cases} \tag{1}$$

Although Cournot's principle has long been fundamental to statistical theory, current philosophical fashion has made it difficult to teach. A frequent objection is that some event of small probability always happens. When we hear this objection, we emphasize "specified in advance", which requires less emphasis in game-theoretic probability, because a bet, by definition, is made in advance.

[5]The anonymous 13th-century author who left us with the earliest surviving calculation of the chances for a throw of three dice warned us (Hexter et al., 2020, p. 172): "Addeque, quod lusor se continuare lucrando nescit, perdendo nescit dimittere ludum." Not knowing how to maintain his luck when winning, the gambler does not know how to quit when losing.

In some cases, we may substitute "simple to describe" for "specified in advance". This also goes without saying in game-theoretic probability, because a bet cannot be made and implemented unless the event is relatively simple.

Cournot's principle can be considered a special case of the fundamental principle for testing, because $1/\alpha$ is the capital that would result from E's happening if you bet unit capital on E.

5 Extending Cournot's principle to test variables

This extension of Cournot's principle does not require us to specify in an advance a goal $1/\alpha$ for the strength of the evidence.

Suppose S is a nonnegative random variable, chosen in advance and so not too hard to describe, with $\mathbb{E}_P(S) = 1$ (call S our *test variable*). Our next principle says that a realized value s of S discredits P to the extent that s is much larger than 1.

Principle 3 (Authorization to test with a test variable)**.** *If we specify a test variable S in advance, then we may take s, the observed value of S, as a measure of evidence against P. We then interpret s (our test score) on the same scale as we use in Cournot's principle. In other words, when $s = 1/\alpha$, it has the same weight against P as the happening of a pre-specified event E when $P(E) = \alpha$.*

Cournot's principle is the special case of Principle 3 where S is given by (1). Principle 3 adds the possibility of a more graduated report on the strength of the evidence against P.[6]

It might seem that the greater flexibility offered by a test variable S comes at a price. When s is the realized value, the events $\{S = s\}$ and $\{S \leq s\}$ happen, and Markov's inequality tells us that our score $1/P(E)$ would have been at least as great, often greater, had we chosen one of these events as our test event E. But of course we could not have made these choices, because we did not know s in advance.

Like the classical form of Cournot's principle, Principle 3 can be considered a special case of the fundamental principle for testing by betting. The observed value s of the test variable S is the capital that would result from buying S for its expected value.

6 Extending Cournot's principle to test martingales

Now suppose we want to test a probability distribution P for a stochastic process $X := X_1, X_2, \ldots$, and we observe the X_t successively. We use a *test martingale*, a nonnegative martingale $S_1, S_2, \ldots$ with $\mathbb{E}_P(S_1) = 1$, again chosen in advance and hence relatively simple. The value s_t of S_t may become known to us only when we have observed $X_1, \ldots, X_t$. To interpret s_t, we adopt this principle:

[6] A more widely used way of obtaining a more graduated report is to use p-values; see §9.

Principle 4 (Authorization to test with a test martingale)**.** *If we specify a test martingale* $S_1, S_2, \ldots$ *in advance, then at all times* t *we may take* s_t*, the observed value of* S_t*, as the current measure of evidence against* P*. If we want, we may stop at time* t *and continue to regard* s_t *as our measure of evidence against* P*. We may interpret each* s_t *(each test score) on the same scale as we use in Principles 2 and 3. In other words, when* $s_t = 1/\alpha$*, it has the same weight against* P *as the happening of a pre-specified event* E *with* $P(E) = \alpha$*.*

Principle 3 is the special case of Principle 4 where P says that all the S_t are equal to each other, so that nothing can be accomplished by continuing past $t = 1$.

Like Cournot's principle and our previous extensions of it, Principle 4 can be considered a special case of the fundamental principle of testing by betting. For each t, s_t is the capital obtained at time t if we first buy S_1 and then at every step invest all our winnings so far in S_t. But we are still testing a mathematical object, a probability distribution P. Forecaster is following a fixed strategy, which tells him to use P's successive conditional probabilities as forecasts, and Skeptic's strategy (the test martingale) is also specified in advance. Neither Forecaster nor Skeptic has a role that allows them to exercise optional continuation. So Principle 4 is not a principle of optional continuation in the sense of this article. It is, however, a principle of optional stopping.

Although the statement of 4 does not mention betting, I do not recall seeing the principle explained or advocated without a betting story.

7 Improvised testing (optional continuation for Skeptic)

Principle 4 authorizes the statistician to use a test martingale specified in advance. Improvisation is not yet authorized. For this, we need some further principle. As with Principle 4, we are testing a probability distribution P for a stochastic process $X := X_1, X_2, \ldots$, and we observe the X_t successively. When $x_1, \ldots, x_{t-1}$ are possible values of $X_1, \ldots, X_{t-1}$, we call a nonnegative variable $S(X_t)$ a *round-t test variable given* $x_1, \ldots, x_{t-1}$ if $\mathbb{E}_P(S(X_t)|x_1, \ldots, x_{t-1}) = 1$; when $t = 1$, this reduces to $\mathbb{E}_P(S(X_1)) = 1$. We can formulate a principle for improvisation in testing as follows:

Principle 5 (Authorization to wing it when testing)**.** *Suppose we set* $s_0 = 1$*, specify a round-1 test variable, say* $S_1(X_1)$*, and then, beginning with* $t = 1$*,*

1. *we observe* X_t*'s value* x_t*,*
2. *we set* $s_t := s_{t-1} S_t(x_t)$*, and*
3. *we specify a round-*$(t+1)$ *test variable given* $x_1, \ldots, x_t$*, say* $S_{t+1}(X_{t+1})$*.*

Suppose we continue for as long as we want and stop whenever we want (after step 2 for some t*). Then at all times* t *until after we stop, we may take* s_t *as the current measure of evidence against* P*. We may interpret* s_t *on the same scale as we use in Principles 2, 3, and 4.*

Principle 5 generalizes Principle 4, and like Principle 4, it can be considered a special case of the fundamental principle for testing by betting. Skeptic is now a free player, not constrained to follow a strategy specified in advance.

8 Improvised probabilities (optional continuation for Forecaster)

Principle 5 authorizes a statistician testing a probability distribution to improvise. But this still does not bring us to R. A. Fisher's vision, where the statistician helps construct over time not only a test but also the probabilities being tested. In this vision, statistician and scientists brainstorm to design an experiment with outcome X_1, to which they assign probabilities based on some theory they want to test, and after observing $X_1 = x_1$, they brainstorm again about what they have learned and design a possibly unanticipated experiment with outcome X_2, and so on.

It is tempting to try to square traditional probability with Barnard's vision by imagining that this collaboration defines a probability distribution P progressively. The first design includes a probability distribution P_1 for X_1. The second includes a probability distribution P_2 for X_2, etc. The product $P_1 \times \cdots \times P_k$, where k is where the research team stops, is a probability distribution P.

But the statistician did not set out to test $P_1 \times \cdots \times P_k$. She and her colleagues waited to design the second experiment and its X_2 and P_2 until they had seen x_1. Had x_1 come out differently, their subsequent brainstorming might have produced a different X_2 and P_2, and so on. If there is a probability distribution being tested, it would seem to involve conditional probabilities for X_2 given all the different x_1 that might be observed (and perhaps also all the other ways the research team's information and thinking might evolve while the first experiment was being performed). And so on.

Some decades ago Dawid (1984, 1991), A. Philip Dawid bravely argued that these dependencies should not matter—that we can design significance tests, confidence intervals, and Bayesian procedures that are unaffected by probabilities, somehow true or somehow invented, involving the might-have-beens. As these might-have-beens do not matter, we can just pretend that we have the requisite independence. This is Dawid's *prequential* model. Although some statisticians (including myself) found it appealing, others found it confusing. What are we really testing? Are we testing a huge and not fully specified probability distribution P whose unspecified probabilities include probabilities for actions of the research team doing the testing?

Leaving all this aside, can we formulate a principle that authorizes us to use Dawid's insight to construct test scores? Here's a try.

Principle 6 (A prequential testing principle)**.** *Suppose we set $s_0 = 1$, construct an experiment that will produce a variable X_1, a probability distribution P_1 for X_1, and a test variable S_1 for P_1, and then, beginning with $t = 1$,*

1. *we observe X_t's value x_t,*
2. *we set $s_t := s_{t-1}S_t(x_t)$, and*

3. *we construct an experiment (perhaps newly conceived) that will yield a variable X_{t+1}, a probability distribution P_{t+1} for X_{t+1}, and a test variable S_{t+1} for P_{t+1}.*

Suppose we continue for as long as we want and stop whenever we want (after step 2 for some t). Then at all times t until after we stop, we may take s_t as the current measure of evidence against the P_t we have constructed so far all being valid. We may interpret s_t on the same scale as we use in Principles 2, 3, 4, and 5.

Principle 5 is a special case of Principle 6. And Principle 6, like our preceding extensions of Cournot's principle, can be considered a special case of the fundamental principle for testing by betting. Now both Forecaster and Skeptic are free agents, not constrained to follow any strategy specified in advance.

The principle's consistency with testing in the game-theoretic framework is not surprising, as that framework was partly inspired by Dawid's prequential model.

9 The role of Ville's inequality.

Ville's inequality says that if $S_1, S_2, \ldots$ is a test martingale, then

$$P\left(\sup_{t\geq 1} S_t \geq \frac{1}{\alpha}\right) \leq \alpha.$$

Some people (including myself) have sometimes said that Ville's inequality authorizes optional continuation. This is a careless formulation. First because a theorem is never more than mathematics; it cannot authorize anything. Secondly because the principle it suggests is not an optional continuation principle developed in this article.

Ville's inequality tells us that $1/\sup_{t\geq 1} S_t$ is a "p-variable" and so $1/\sup_{t\geq 1} s_t$ is a p-value. Well, almost. It is at least implicit in the notion of a p-value, as statisticians understand and use the term, that we have observed it and know we have observed it. We do not expect this for $1/\sup_{t\geq 1} s_t$. But we do observe upper bounds. At time t, we have observed the upper bound $1/\sup_{1\leq i\leq t} s_i$, and an upper bound on a p-value is a p-value. So most statisticians who use p-values would probably accept this principle:

Principle 7 (The dynamic p-value principle)**.** *As we continue to make observations, we may always use the current* $1/\sup_{1\leq i\leq t} s_i$ *just as statisticians usually use a p-value.*

This principle is implicit in the use of confidence sequences, which go back to Darling and Robbins (1967).

Principle 7 can be considered an optional stopping principle, because it authorizes us to use $1/\sup_{1\leq i\leq t} s_i$ like a p-value if we stop at time t. But it is more than an optional stopping principle, because it authorizes us both to use $1/\sup_{1\leq i\leq t} s_i$ like a p-value at time t *and also* to continue. It is not an optional continuation principle in the sense of this article, because it does not authorize us to change the later experiment (the probabilities for future Xs) or the test martingale.

As an optional stopping principle, Principle 7 can be compared with Principle 4. Neither is stronger than the other. Principle 4 authorizes us to use s_t as a measure of our evidence against P and to continue doing so if we stop. But it does not allow us to continue using s_t if we do not stop and hence does not authorize us to use the sometimes larger $\sup_{1 \leq i \leq t} s_i$ (Shafer et al., 2011). But it gives $1/s_t$ the force of a fixed significance level, which is greater than the force of a p-value.

Ville's inequality and Principle 7 have generalizations in game-theoretic statistics, where they use game-theoretic definitions of upper and lower probability and expected value. See (Shafer and Vovk, 2019, Exercise 2.10).

10 Conclusion

We have shown that the traditional principle for testing a probability distribution (Cournot's principle) can be extended so that it fully accommodates optional continuation and yet does not explicitly use game-theoretic probability or ideas about betting. Is this extension worth the trouble?

The clear message of the exercise is that the fundamental principle of testing by betting, coupled with game-theoretic probability, provides a theoretical basis for optional continuation that is simpler, clearer, and more general. Readers will judge for themselves, but I submit that Principle 6 is overly complex, ill-motivated, and impossible to teach without reference to betting. It remains, moreover, less general than the fundamental principle of testing by betting, because it requires Forecaster's moves on each round of a forecasting game to be a probability distribution.

Our exercise has also illustrated the new clarity brought to statistical theory by game-theoretic probability's distinction between Forecaster and Skeptic. This distinction has helped us see the complexity of the notion of optional continuation. Optional continuation for Forecaster is a step further than optional continuation for Skeptic.

11 Acknowledgments

In recent years I have been able to discuss this topic with a number of colleagues, especially Gert de Cooman, Peter Grünwald, Aaditya Ramdas, Judith ter Schure, and Volodya Vovk. I have had especially useful and stimulating conversations at the recent workshop *Safe, Anytime-Valid Inference (SAVI) and Game-theoretic Statistics* (May 30–June 3, 2022 in Eindhoven, Netherlands) and subsequently while working on a review article with Aaditya, Peter, and Volodya.

This meditation is a product of these discussions, and especially of Peter's insistence on conceptual foundations and Aaditya's refusal to accept more facile

arguments. It has been improved by comments on an earlier draft by Volodya and by Philp Dawid.

References

George A. Barnard. Review of *Sequential Analysis*, by Abraham Wald. *Journal of the American Statistical Association*, 42(240):658–665, 1947. 2

George A. Barnard. Statistical inference. *Journal of the Royal Statistical Society B*, 11(2):115–149, 1949. 2

Donald A. Darling and Herbert Robbins. Confidence sequences for mean, variance, and median. *Proceedings of the National Academy of Sciences of the United States of America*, 58(1):66–68, 1967. 9

A. Philip Dawid. Present position and potential developments: Some personal views statistical theory the prequential approach. *Journal of the Royal Statistical Society: Series A*, 147(2):278–290, 1984. 8

A. Philip Dawid. Fisherian inference in likelihood and prequential frames of reference. *Journal of the Royal Statistical Society: Series B*, 53(1):79–109, 1991. 8

Persi Diaconis and Brian Skyrms. *Ten Great Ideas about Chance*. Princeton, 2018. 1

Joseph L. Doob. *Stochastic Processes*. Wiley, New York, 1953. 2

William Feller. *An Introduction to Probability Theory and Its Applications*. Wiley, New York, 1st edition, 1950. 2

Willy K. Feller. Statistical aspects of ESP. *The Journal of Parapsychology*, 4 (2):271–298, 1940. 2

Ronald A. Fisher. Sequential experimentation. *Biometrics*, 8:183–187, 1952. 2

Joseph Albert Greenwood. An empirical investigation of some sampling problems. *The Journal of Parapsychology*, 3(2):222–230, 1938. 2

Peter Grünwald, Alexander Ly, Muriel Pérez-Ortiz, and Judith ter Schure. The safe logrank test: Error control under optional stopping, continuation and prior misspecification. *Proceedings of Machine Learning Research*, 146:107–117, 2021. 2

Peter Grünwald, Rianne De Heide, and Wouter Koolen. Safe testing, 2023. arXiv:1906.07801; to appear in *Journal of the Royal Statistical Society B*. 2

Allard A. Hendriksen. Betting as an alternative to p-values, 2017. Master's thesis, University of Leiden, under the direction of Peter Grünwald. 2

Ralph Hexter, Laura Pfuntner, and Justin Haynes, editors. *Appendix Ovidiana: Latin Poems Ascribed to Ovid in the Middle Ages*. Harvard, 2020. 5

Glenn Shafer. From Cournot's principle to market efficiency. In Jean-Philippe Touffut, editor, *Augustin Cournot: Modelling Economics*, pages 55–95. Edward Elgar, 2007. See also Working Paper 15, Game-Theoretic Probability Project. 1

Glenn Shafer. Game-theoretic foundations for statistical testing and imprecise probabilities, December 2020. Slides and videos. 4

Glenn Shafer. That's what all the old guys said: The many faces of Cournot's principle, 2022. Working Paper 60, Game-Theoretic Probability Project. 1

Glenn Shafer and Vladimir Vovk. *Probability and Finance: It's Only a Game*. Wiley, New York, 2001. 1, 3

Glenn Shafer and Vladimir Vovk. The sources of Kolmogorov's *Grundbegriffe*. *Statistical Science*, 21(1):70–98, 2006. See also Working Paper 4, Game-Theoretic Probability Project. 1

Glenn Shafer and Vladimir Vovk. *Game-Theoretic Foundations for Probability and Finance*. Wiley, New York, 2019. 1, 3, 10

Glenn Shafer, Alexander Shen, Nikolai Vereshchagin, and Vladimir Vovk. Test martingales, Bayes factors and p-values. *Statistical Science*, 26(1):84–101, 2011. 10

Abraham Wald. *Sequential Analysis*. Wiley, New York, 1947. 2