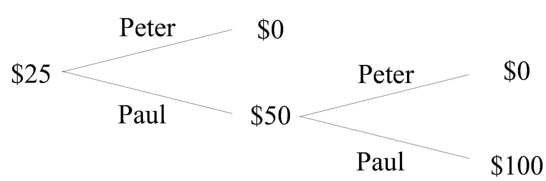
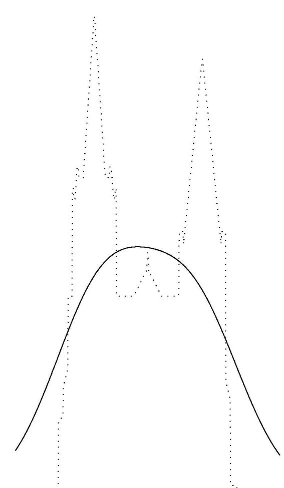
A meditation on optional continuation

Glenn Shafer
Rutgers University
gshafer@business.rutgers.edu
www.glennshafer.com





The Game-Theoretic Probability and Finance Project

Working Paper #63

First posted August 4, 2022. Last revised August 4, 2022.

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

Abstract

How is optional continuation justified in game-theoretic statistics? Does game theory provide a justification not available in measure theory?

In this note, I try to answer these questions for myself. Others may already understand everything I say. But I try to say it clearly enough that I can understand it myself and perhaps remember it.

Contents

Optional continuation, as I am using the term, is a property attributed to some statistical methods, including the method of testing by multiplying likelihood ratios from successive observational studies or experiments. When we say that a statistical method permits optional continuation, we are saying that a statistician using the method is authorized to decide whether to continue collecting data and even whether to change the experiment after seeing outcomes so far. These decisions need not be planned in advance. They need not follow a strategy adopted at the outset. Yet they do not invalidate measures of evidence that the method provides.

I first noticed "optional continuation" being used with this meaning a few years ago, in an early version of [9].[1] But the idea is much older. It goes back to the introduction of the term "optional stopping" by the Duke mathematician Joseph Albert Greenwood in 1938 [8, pp. 225–226]. Greenwood tried to understand empirically the adjustments needed to account for the way Joseph Rhine's laboratory was opportunistically switching participants in its experiments on extra-sensory perception. When a participant was no longer succeeding at their task, the experimenters would *stop* using that participant and *continue* with someone fresh.

Greenwood's "optional stopping" was brought to wider attention in mathematical statistics by William Feller in a 1940 critique of the ESP work [7, pp. 272, 286–292] and in the first edition of his textbook on probability, which appeared in 1950 [6, pp. 140, 190, 197]. As subsequently used in mathematical probability, especially by Joseph Doob [5], "optional stopping" seems to be about plans for stopping (stopping rules). Peter Grünwald and his collaborators are now using "optional continuation" in the broader way Greenwood originally used "optional stopping", where the experimenter may even continue with a different experiment.

In his 1947 book on sequential analysis [17], Abraham Wald considered only "sequential sampling plans" chosen in advance. But in a review of the book that same year [1], George Barnard wrote that sequential analysis marked "the entry of statistical considerations into the very process of experimentation itself." As now seems obvious, the process of experimentation often involves not only decisions about whether to stop but also decisions about how to change the sampling program or experiment.

1 Can mathematics authorize optional continuation?

Well, when you put it that way, we have to say no. Mathematics may be able to tell us what is true, but it cannot tell us what to do. To go from a bit of mathematics to concluding something about the world or to deciding to do

[1] The earliest such use I have found in Google Scholar is in Allard Hendriksen's 2017 master's thesis at the University of Leiden [10]. On page 3, Hendriksen writes,

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

As of July 29, 2022, "optional continuation" had not yet appeared in any of the 34 statistics journals in JSTOR.

something in the world, we must add something more. Extra-mathematical assumptions or principles of some sort.

What principles do we need to add to measure-theoretic probability or to game-theoretic probability in order to authorize optional continuation?

2 Game theory's marginal standing in mathematics

Before calling our two contenders, measure theory and game theory, into the ring, let's think about each's advantages and disadvantages.

Measure theory's great advantage is its status within mathematics. In my youth, game theory rode in the back of second class of the mathematics train, perhaps even behind statistics. Measure theory, on the other hand, consorted with analysis. If number theory was still the queen of mathematics, analysis was the king. Perhaps this has changed a little. Perhaps some game theorists disguised as economists or computer scientists have infiltrated first class. I do sense that they are in front of the statisticians. Meanwhile, measure theory and even set theory may have become a tad boring. But *grosso modo*, measure theory is still far closer to the center of mathematics than game theory.

The problem with game theory, as it was developed by Émile Borel and John von Neumann in the 1920s, is that it seems to be about people, not about mathematics. It is about people playing a game. To make mathematics of the game, you must filter the people out. You make their possible strategies into formal objects in set theory and prove theorems about these formal objects, ignoring the possibility that the people might play as they like, deciding on each round what to do next without regard to any strategy. Mathematicians will acknowledge, if grudgingly, that theorems about strategies are mathematics. But the smell of the people lingers.

Yet the connection with human action that puts game theory at a disadvantage with measure theory in the competition for status in mathematics may be an advantage when we formulate principles for what statisticians may do.

3 A betting game with optional continuation

The basic game for statistical testing has three players: Forecaster makes probability predictions, Skeptic bets against them, and Reality announces the outcomes. In our 2001 and 2019 books, [14, 15], Vladimir 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.[2] 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. But, again, we can vary the rules, imposing strategies on some of the players (pure strategies, not mixed strategies if we are remaining purely game-theoretic) or adding other players whose moves constrain their choices.

Clearly, a principle needs to be added for the game to be used outside mathematics, where we put you (or some other person or a theory, etc.) in the role of Skeptic. Vovk and I have named this principle in various ways and expressed it in various ways in our books and on other occasions.[3] At the moment, my favorite formulation is the one I used in my SIPTA lectures in December 2020 [12]:

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.*

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 forecasters 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.

[2] To see how Forecaster's flexibility concerning what he forecasts can be built into the game formally in a way that standard probability theorems can still be proven, see [15, §7.5].

[3] In [14], we called the principle the fundamental interpretative hypothesis of probability. In [15], we called it the game-theoretic version of Cournot's principle.

- 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.

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 get the money if you keep betting and lose it.[4]

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 (or in the stock market), you are not much of an expert about it. Etc.

4 Cournot's principle for testing a probability measure

What principles must we add to measure-theoretic probability to allow optional continuation?

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

[4] The anonymous 13th-century author who left us with the earliest surviving calculation of the chances for a throw of three dice warned us [11, 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.

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*:

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

I have documented how pervasive Cournot's principle has been historically (in [13], for example), but it is difficult to teach, as it brings to mind so readily the objection that some event of small probability always happens. When we hear this objection, we emphasize the "specified in advance", which required less emphasis in the game-theoretic story, because a bet, by definition, is made in advance.

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

5 A more flexible version of Cournot's principle

There is a version of Cournot's principle that 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). What Principle 3 adds is the possibility, in many cases, of a more graduated report on the strength of the evidence against P.[5]

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.

[5] A more widely used way of adding this possibility is to use p-values; see §10.

6 Cournot's principle for a stochastic process

Now suppose we want to test a probability measure 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 becomes known to us only when we have observed $X_1, \ldots, X_t$. To interpret s_t, we adopt this principle:

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. We may interpret s_t (our 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 4 may be called an *optional continuation principle*, because we are not obliged to pay any attention to what happens after any particular time. Principles 2 and 3 can be thought of as special cases where P says that all the S_t are equal to each other, so that nothing can be accomplished by continuing past $t = 1$.

We may use betting stories when we teach Principles 2, 3, and 4 and try to persuade people to adopt them. I do not recall seeing 3 and 4 taught without such betting stories. But we are still testing a mathematical object, a probability measure P, not a forecaster in a betting game.

7 Improvised testing

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 measure 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 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.

How would we explain this principle to an applied statistician or an undergraduate student without using the language of betting?

8 Improvised probability measure?

Principle 5 authorizes a statistician testing a probability measure to improvise. But this still does not bring us to George Barnard's vision, where the statistician helps construct over time not only a test but also the probabilities being tested. In this vision, the statistician and the 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 measure-theoretic probability with Barnard's vision by imagining that this collaboration defines a probability measure P progressively. The first design includes a probability measure P_1 for X_1. The second includes a probability measure 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 measure 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 measure 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 [3, 4], 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, it has not been widely implemented, perhaps because the mathematical overhead is unclear. Are all statistical results obtained using $P_1 \times \cdots \times P_k$ really consistent with results that might be obtained testing other P in which the P_t appear as conditional probabilities, and what would be involved in proving this with measure-theoretic rigor? Alternatively, is it mathematically and philosophically coherent to say that we are testing a huge and not fully specified probability measure 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 yield 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 that will yield 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.

We can think of the setup here as an example of the betting game described in §3, and then Principle 6 becomes an instantiation of the fundamental principle for testing by betting. The presence of probability measures on each round differs not at all from their presence in many other betting games studied in [15]. This is not surprising, because the game-theoretic framework developed in [15] was largely inspired by Dawid's prequential model.

9 What's the verdict?

In the first paragraph of this note, I asked whether game theory provides a justification of optional continuation not available in measure theory. The reader will agree that neither game theory nor measure theory, considered simply as mathematics, provides such a justification. So the question is which theory best accommodates added principles that authorize optional continuation. The reader will render their own verdict on this.

My own verdict is that game theory emerges as the winner. The fundamental principle for testing by betting is relatively brief and coherent, expressible in English in 26 words. When we contort measure theory to authorize optional continuation in its fullest sense (where we construct both probabilities and tests as we proceed), the resulting principle, Principle 6, is essentially an instantiation of this fundamental game-theoretic principle.

10 The role of Ville's inequality.

Ville's inequality says that if $S_1, S_2, \dots$ 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, because a theorem is never more than mathematics; it cannot authorize anything.

Ville's inequality does tells us something relevant to optional continuation. It 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 1967 [2].

Like Principle 4, Principle 7 is an *optional continuation principle*. But the two principles are not the same, and neither is stronger than the other. Principle 4 authorizes us to use s_t; it does not authorize us to use the sometimes larger $\sup_{1\leq i\leq t} s_i$. 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 are also available in game-theoretic statistics, where they use game-theoretic definitions of upper and lower probability and expected value. See [15, Exercise 2.10].

11 Why weren't we already talking about this in the 1950s?

Abraham Wald (1902–1950), Joseph Albert Greenwood (1906–1988), and George Barnard (1915–2002) came to age well before the Andrei Kolmogorov's axioms and Joseph Doob's filtrations were widely accepted as *the* foundation for mathematical probability and statistics. In a 1942 a lecture on the foundations of probability [16], Wald mentioned von Mises's axioms and Kolmogorov's axioms as equally viable foundations. Doob's classic *Stochastic Processes*, which showed mathematicians how to understand martingales in terms of filtrations, did not appear until 1953 [5]. So we cannot be surprised that Wald and Barnard did not try to fit their probabilistic intuitions into the Kolmogorov/Doob framework in 1947. Perhaps it is more surprising that later statisticians have devoted so little attention to the difficulty of doing so.

Or perhaps not. After Wald's death and the rallying of mathematicians to Doob's framework, it became axiomatic in mathematical statistics that acceptable statistical theory must at least pay lip service to that framework. What could not be said within it could not be said at all.

12 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

[1] George A. Barnard. Review of *Sequential Analysis*, by Abraham Wald. *Journal of the American Statistical Association*, 42(240):658–665, 1947. 1

[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

[3] 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. 7

[4] 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. 7

[5] Joseph L. Doob. *Stochastic Processes*. Wiley, New York, 1953. 1, 9

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

[7] Willy K. Feller. Statistical aspects of ESP. *The Journal of Parapsychology*, 4(2):271–298, 1940. 1

[8] Joseph Albert Greenwood. An empirical investigation of some sampling problems. *The Journal of Parapsychology*, 3(2):222–230, 1938. 1

[9] Peter Grünwald, Rianne De Heide, and Wouter Koolen. Safe testing, 2019. arXiv:1906.07801; accepted pending minor modifications by the *Journal of the Royal Statistical Society, series B*. 1

[10] Allard A. Hendriksen. Betting as an alternative to p-values, 2017. Master's thesis, University of Leiden, under the direction of Peter Grünwald. 1

[11] Ralph Hexter, Laura Pfuntner, and Justin Haynes, editors. *Appendix Ovidiana: Latin Poems Ascribed to Ovid in the Middle Ages*. Harvard, 2020. 4

[12] Glenn Shafer. Game-theoretic foundations for statistical testing and imprecise probabilities, December 2020. Slides and videos. 3

[13] 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. 5

[14] Glenn Shafer and Vladimir Vovk. *Probability and Finance: It's Only a Game*. Wiley, New York, 2001. 2, 3

[15] Glenn Shafer and Vladimir Vovk. *Game-Theoretic Foundations for Probability and Finance*. Wiley, New York, 2019. 2, 3, 8, 9

[16] Abraham Wald. *On the principles of statistical inference*. University of Notre Dame, 1942. 9

[17] Abraham Wald. *Sequential Analysis*. Wiley, New York, 1947. 1