Cambridge University Press
978-1-107-02415-1 - Measuring and Reasoning: Numerical Inference in the Sciences
Fred L. Bookstein
Excerpt
More information

Part I

The Basic Structure of a Numerical Inference

1

Getting Started

1.1 Our Central Problem: What Is a Good Numerical Inference?

You know this move. You have seen it dozens of times, or thousands if you have been a scientist or professor for long enough. A text or a PowerPoint slide presents some arithmetic computation based on measured data – group averages, or a regression slope, or the range of some measurement, or a bar chart or time series – and then the author continues, "So I have shown . . ." or, more conspiratorially, "We have thus shown . . ." or, more didactically, "It follows that . . ." where the ". . ." in all cases is some proposition more general or otherwise more assertive than the scope of the data actually reported. The qualitative assertion may deal with a cause, or a consequence, or a generalization about past, hypothetical present, or future under some eutopian or dystopian policy. Or the sentence may be a bit more sophisticated, with the same import: "Hence our null hypothesis is rejected" or, the form with the most internal evidence of thoughtfulness, "Hence Professor Smith's theory is false."

Here are a few examples unsystematically extracted and rephrased as single sentences from later in this book. "It followed that the submarine was traveling east." "The two continental plates are moving apart at a rate of about four centimeters per year." "Cholera is caused by some morbid material in your drinking water." "There was an anthrax epidemic in Sverdlovsk, Russia, in 1979, that originated in a secret biowarfare factory." "Einstein's law of the photoelectric effect is correct." "Atoms exist." "Stomach ulcers are caused by a bacillus, *Helicobacter pylori*." "The form of DNA is a double helix pairing adenine with thymine and guanine with cytosine." "Environmental tobacco smoke raises your chances of a heart attack or of lung cancer by just about one-fourth even after considering differences in diet." "Civic worth is inherited just as much as eye color." "This is the skull of Oliver Cromwell." "This brain was damaged before birth by its exposure to high levels of alcohol."

There are many more general tropes having this form. It is fun to estimate the propensity of stereotyped research summaries in this family by using a Google search in its "sentence completion" mode. If you type "Boys have more" (including the quotation marks) into Google, a search returns more than 6 million pages with this phrase. On one Tuesday morning in 2013, the first page of retrievals includes "Boys have more health problems in childhood than girls" (a study from *Acta Paediatrica* of Finnish children born in 1987); "Women expecting boys have more testosterone" (and therefore weaker immune systems, hence shorter life spans), ostensibly a health story from the *Daily Mail*, a British newspaper; "Boys have more cortical area devoted to spatial-mechanical functioning," from a University of Arizona keynote address about gender and mathematics achievement; and "Why do boys have more freedom than girls?," a finding that likely depends on details of the way the term "freedom" was operationalized. The completion template "Apes have less" retrieves almost 240,000 hits; "People with a high IQ," 410,000. It may be that this sort of binary reclassification semantic (more or less, higher or lower, no matter how much) is how computations get turned into propositions for the majority of contexts in the social and medical sciences, in popular wisdom, and in real life in the modern Western setting.

But how does this rhetorical and semantic process actually operate? How do we readers or listeners move from one sentence to the next across this divide, from displays of arithmetic to the cognitive focus on some inference? I am not questioning (at least, not in this book) the arithmetic, nor the fact that the numbers entering into the arithmetic arose from some sort of instrument reading that the experimenter could distort more or less at her own volition. Rather, I am questioning this central trope, usually unspoken, that guides us over the leap from arithmetic to inference – from routinized observations out of some empirical data set, with all of its accidents and contingencies, into an intentionally more abstract language of a truth claim. After this maneuver, the conversation has ceased to deal with "formulas" or "statistical methods" or other merely arithmetical concerns, and has now become a matter of qualitative assertions and their implications for human beliefs or human actions.

This book is concerned with the specific process by which a numerical statement about a particular data set is transformed into a more general proposition. That is the process I am calling "numerical inference." It is not syllogistic (though we see an attempt to make it so in the discussion of C. S. Peirce's ideas in Chapter 3). Cognitively, it would seem to be instantaneous, rather like the "Eureka!" moment reported by James Watson in his narrative of the discovery of the double helix (see at "Suddenly I became aware . . . ," Section E4.5).[1] The inference itself seems intuitive. Epistemological discussions such as those of this book at best supply justifications after the fact or the mental discipline that renders that justification easier on a routine basis.

The basic inference itself – the claim that an essentially verbal proposition follows from an arithmetical manipulation of some measurements – is not intrinsically fallacious. Sometimes support for a proposition *does* follow from data – for instance, most of the examples in Chapter 4 are valid. But in almost all cases the inference went unjustified *in the course of the text in which it appears.* (For superb exceptions, see Perrin's argument for believing that atoms exist, Section E4.3, and also my last

[1] Throughout this book, crossreferences to extents less than an entire chapter are indicated by the word "Section" followed by a subsection or sub-subsection specification, such as "Section 5.3.1." In Chapter 4, after the brief introduction, subsections are further sorted as "lectures" with a heading beginning with "L4" or instead "examples" with headings beginning "E4." So crossreferences to this chapter include that letter, for instance, "Section L4.3.1" or "Section E4.3."

Chapter 4 exemplar, the argument by Alvarez et al. about one big reason for the extinction of the dinosaurs.)

There are ways of designing one's studies so that the numerical inference is more likely or less likely to be valid, and discussions of these ways make up much of the content of this book. But first it is necessary to foreground these conventionally tacit links inside conventional scientific narratives – to bring them out from the dark into the light so that their typical features can be noted and critiqued. Neither the natural sciences nor the social sciences could operate effectively without them. Our task is not to eliminate them, but to highlight them: first, to detect them when they appear, or to point out their omission at places where they should appear but do not; second, to suggest a prophylaxis by which their average cogency and reliability might be improved.

In other words: the central concern of this book is not for the *content* of a scientific argument but for its *flow* in an important special case, to wit, the rules and the logic of inferences from numbers or numerical measurements. The question is not "How do you state a conclusion once you have analyzed your numerical data?" For that there are a great many books at a great many levels of sophistication, of which the oldest in my collection is Huff's 1954 *How to Lie With Statistics*. Rather, the issue is under what circumstances such an inference can be trustworthy. *When* do inferences from numbers make scientific sense, and how?

Central domains for my examples here include the physical and biological/biomedical natural sciences and also the historical/biographical natural and social sciences. What these fields share is a respect for consilience together with the possibility of abduction. **Consilience** (here I anticipate the long introduction in Chapter 2) is the matching of evidence from data at different levels of observation or using essentially different kinds of measurement; **abduction**, the main topic of Chapter 3, is the sequence of finding something surprising in your data and then positing a hypothesis whose specific task is to remove that surprise. **Strong inference**, a sort of multiple abduction in parallel, is the wielding of a single data set to support one hypothesis at the same time it rejects several others that had previously been considered just as plausible.

But this has all been rather abstract. Before going any further, let us dive right into a particularly rich example from an unexpected quarter, namely, the ocean floor about 400 miles northwest of the Azores.

1.2 The Sinking of the *Scorpion*

The *Scorpion*, a submarine of the United States Navy, sank on May 21, 1968, with loss of all hands. I have no independent knowledge of this event, but proceed by deconstructing two easily accessed readings on the topic: the chapter "Death of a Submarine" in Sontag and Drew (1998), and the Wikipedia page `http://en.wikipedia.org/wiki/USS_Scorpion_(SSN-589)`. For an alternative overview of this same episode, see McGrayne (2011, pp. 196–204).

The basic structure of the investigation here is easy to summarize. A submarine vanished, and the U.S. Navy wished to learn why so that similar accidents to her sister ships might be avoided. Although the data pertain to one historical event only, the inference is a generalization – to levels of risk across a class of vessels. What makes the case particularly interesting is that to begin with there were no data – those had to be constructed (not reconstructed) before any inferences could be grounded at all. A cycling was needed between potential inferences and the data resources that might render them plausible, thereby motivating further potential inferences suggesting the accrual of more new evidence in turn.

The story begins with a bald fact: an American submarine, the *Scorpion*, didn't show up at its base in Norfolk, Virginia, on May 27, 1968. The fact arrives unconnected to any narrative that could give it meaning. It has no associated numerical characteristics, but is merely a statement about the absence of something.

One has to begin somewhere, and the argument in this instance begins with the declared plan communicated by the submarine to its base prior to this journey. Its intended travel was along a fixed path, and we can assume (as did the Navy) that its own instruments were adequate to keep it on this path: a great circle (the shortest possible route) to be traversed underwater at a carefully monitored speed of 18 knots (about 20 miles per hour) all the way from the Azores to Norfolk. This intention counts as a preliminary "theory" in the discussion that follows: an explanatory structure, evidence-driven deviations from which remain to be discovered.

The *Scorpion* could have been anywhere on this 3000-mile track when something untoward happened. The task is to figure out where it was, and then *either* to account for what might have happened to it *or* to go take a look at the wreckage to see what causes of catastrophe might be inferrable from the image. Historically, this inference was in the hands of one man, Navy analyst John Craven, and it is in Craven's voice that Sontag and Drew tell the story. (For Craven's own retelling, see Craven, 2002.)

At this point a glimmer appears of the first of a series of inferences whose logical machinery concerns us in more detail in Chapters 2 and 3. The "track" is actually a path in space and time, a particular pairing of location with time that could perhaps be matched to instrument readings to produce evidence of particularly interesting moments that might engender a possible explanation or alter an explanation that had been proffered earlier. There is a tacit model here about the machine that was the submarine itself – that it was functioning within specifications (i.e., staying on that space-time track) until the moment of some disaster – and there are shortly to emerge some assumptions about other machines, namely, the listening devices that will highlight specific moments along this planned trajectory.

For our purposes, the main logical characteristic of machines is the predictability of their performance as long as they are "working." This presumption of predictability applies both to active machines like submarines and to the passive devices that make observations. That predictability is one of the deepest cruces in all of natural science, closely tied to the principle of consilience on which this book is ultimately based. As Wigner (1960, p. 5) put it, "the construction of machines, the functioning of which he can foresee, constitutes the most spectacular accomplishment of the physicist. In these machines, the physicist creates a situation in which all the relevant coordinates

Cambridge University Press
978-1-107-02415-1 - Measuring and Reasoning: Numerical Inference in the Sciences
Fred L. Bookstein
Excerpt
More information

are known so that the behavior of the machine can be predicted. Radars and nuclear reactors are examples of such machines." And, I would add, attack submarines.

Along with this theory of submarine behavior – that in the absence of any other factor they tend to travel along great-circle routes – are additional facts that seem, but only seem, to be numerical in the same way. The *Scorpion*, for instance, was restricted to depths of less than 300 feet, owing to delays in some safety refittings; but this value of 300 will not enter into any inferences, it turns out. The sub had experienced a bad bout of wild accelerations and vibration some months earlier – a "corkscrewing problem" – but the specific physical parameters of that episode likewise will play no role in the inferences to come. Their irrelevance could not be asserted, though, until a valid theory had finally been asserted and verified. We are illustrating in passing a point from a discussion in Section 2.5 that much of the art of physics inheres in knowing what to measure and what to ignore.

A first step along the road to discovery was the simple scan of historical records to rule out collisions or battles. In the language of Section L4.6, these are "plausible rival hypotheses," easily ruled out by searches through databases whose details do not concern us here.

The numerical inference begins to take shape with the search for instrument records that might be informative. (There is a close analogy here to the reliance of evolutionary demography on massive databases, such as the Integrated Public Use Microdata Series [IPUMS], that have been accumulated by bureaucracies for other purposes entirely. See Huber et al., 2010.) One set, the Navy's own SOSUS (Sound Surveillance System), was systematically useless for Craven's purpose, by design: it was set to filter out all point events (like blasts) in order to highlight the periodic noises of Soviet machines. But an Office of Naval Research laboratory in the Canary Islands happened to have preserved a set of unfiltered records for the week in question, and within them were eight separate bursts of signals at amplitude well above background. Working backwards from time to place, this gave the Navy eight different locations to check for surface wreckage. Surveillance planes found nothing, or else there would be no story to tell here.

Now this is not a consilience yet (to anticipate the language of Chapter 2). Any event in the Canary Islands record for that week must match *some* location on the *Scorpion*'s track, and so before any grounded inference could begin there needed to be a coincidence (in the literal sense, co-incidence, meaning co-occurrence) between two channels of data, not just a map line and a microphone reading separately. Craven needed *another* source of information, as independent as possible. (We go more deeply into this notion of "independence" of information resources in Section 2.3, in connection with Collins's discussion of replications.) There were two more hydrophones in the Atlantic, intended to track Soviet nuclear tests: one off the peninsula of Argentia in Newfoundland, the other 200 miles away. This pair of positions was not entirely suitable – there were undersea mountains between them and the Azores – but when Craven laid the recordings from the two hydrophones over the Canary recordings, some of the blips lined up. The geodetic situation is approximately as rendered in Figure 1.1.

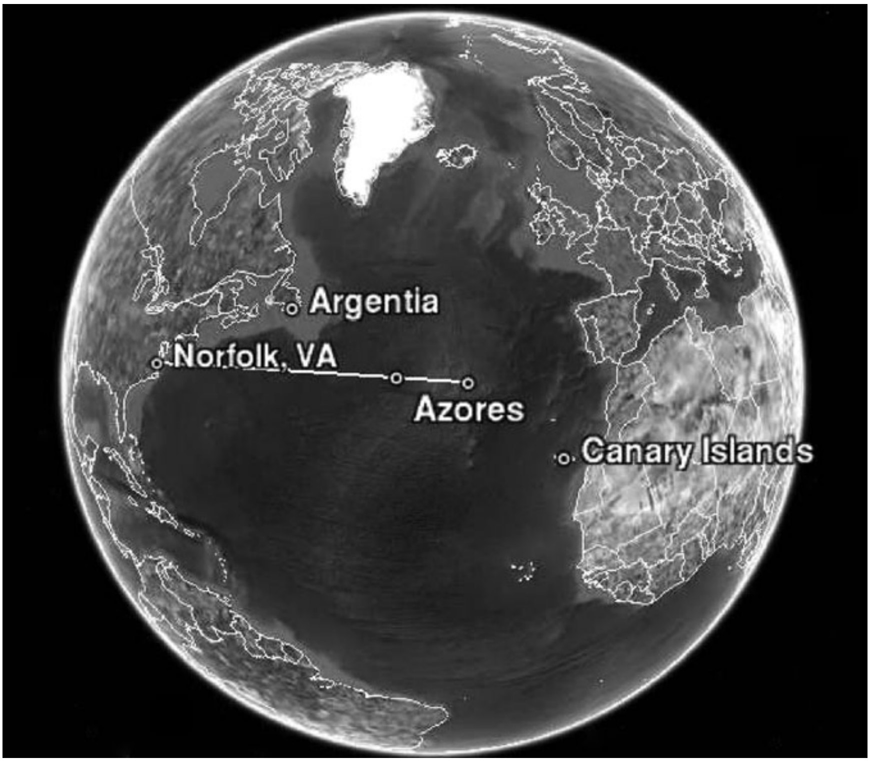

Figure 1.1. North Atlantic locations relevant to the *Scorpion* story. (A GoogleEarth figure from a viewpoint about 8000 miles above the water.) *Scorpion*'s planned track from the Azores to Norfolk, Virginia is the white line shown. The unnamed open circle along this line is the approximate location of "Point Oscar," the point on the track corresponding to the match of interesting loud sounds among the three microphones, one in the Canary Islands and two others in the vicinity of Argentia, Newfoundland.

At this point, there is an actual formal inference, of the style we review at length in Chapter 3. Here it is in Sontag and Drew's words:

> If the Argentia blips were worthless noise, then the plots would probably fall hundreds of miles or thousands of miles from the relatively tiny line of ocean that made up *Scorpion*'s track. But if the new data pinpointed *[sic]* any one of the eight events picked up in the Canary Islands on that tiny line, the acoustic matches would almost certainly have to be valid. (Sontag and Drew, 1998, p. 96)

And they matched, at one first moment and then 91 seconds later, 4 seconds after that, 5 seconds after that, and so on for 3 minutes and 10 seconds. Then silence.

The *inference*, then, is that these sounds recorded the destruction of the *Scorpion* from implosion. It is a qualitative inference from quantitative data – the very topic on which this book's argument is centered. This is the first inference: that these sounds are the sounds of *Scorpion*'s destruction at a particular place and a particular time.

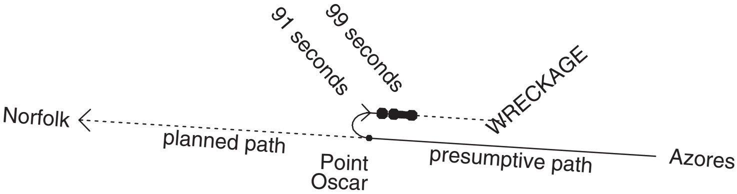

Figure 1.2. Schematic of Craven's understanding after he discovered that *Scorpion* was traveling eastward, not westward, over the series of detected blasts. On the diagram, 91 seconds is the interval between the first detected event and the second, much louder event, and 99 seconds is the interval between the second and the last of the series.

Thus the inference, a particularly powerful one in view of the *repeated* co-occurrences, but it must be validated. To verify the Alvarez argument we review in Section E4.8, somebody had to find a meteor crater of the appropriate size and age; to verify Craven's inference, somebody has to find the wreckage of the submarine. What to do? The submarine could be anywhere in a circle of radius roughly 10 miles around this point of the last implosion signal.

At this juncture three distinct numerical *enhancements* appear, each individually quite brilliant: a sharpening of the information from the instrumental record by filtering, and two separate sharpenings of the theory by human interview. The sharpening of the record is not described in my sources – I am assuming it represented a computation by differentials in the arrival times of the blast signals at the sites of what are by now three microphones (Canary Islands, Argentia, and the third one). From this Craven learned that *Scorpion* was moving *east*, not west, between explosions. Remember that its planned trajectory was *west*, directly west, from the Azores to Norfolk. *Trusting the instruments*, then, he asked what could make a submarine carry out a 180° turn in the middle of the ocean, the scenario diagrammed in Figure 1.2.

The universal answer, according to Craven's sources (mainly retired submarine commanders), was that this maneuver would be the result of a "hot torpedo," the accidental launch of a torpedo or the jettisoning of a defective and dangerous torpedo when there is no target in the vicinity except the submarine itself. It was a standard ploy (a trope, in the language of the torpedo's own guidance system) that instructs the torpedo not to pursue its own mother ship. This then constitutes a change in the theory (here, in the expected trajectory) based on new data.

But this is the wrong theory – it doesn't involve the sinking of a submarine. What else needs to have happened? Craven's sources suspected the "cook-off" of a torpedo – a fire in the torpedo room starting from a bad battery – followed by an explosion there that blew the hatches open, destroying the submarine's buoyancy. Simulations with humans showed that about 90 seconds after the simulated explosion, no matter what a commander did, the submarine would arrive at its implosion depth of about 2000 feet and the series of closely spaced blasts would begin. Recall that the actual interval between the first and the second blasts, from the hydrophone records, was 91 seconds.

This is the second inference: an actual historical sequence of events corresponding to the sequence of instrumented records. Craven now had to shift to a wholly quantitative *third* inference: given this scenario, where, exactly, would *Scorpion* come to rest on the seabed? This takes the form of a classic problem in prediction under uncertainty, along with an associated search strategy that was itself the subject of a book (Stone, 1975). The method for this was not remotely objective. A set of parameters was listed (the wreck's downward glide speed, the direction of glide, the glide slope, and so on). A bottle of Scotch whisky would be the reward for the expert whose predictions would best match the actual location of the sub when it was found. There resulted a prediction of the best (optimal) location to look and a corresponding search strategy. (We discuss the notion of "optimal" estimates of statistical parameters in Section 5.3.)

That was the third inference, leading to new data conveying its information most convincingly indeed. The wreckage of *Scorpion* was found on October 29 within 200 meters of where Craven and the experts had estimated its likeliest location. She lay at a depth of 11,000 feet (3.35 kilometers, just over 2 miles). We are not told who won the bottle of Chivas Regal.

This not only closes the inference about where *Scorpion* was – that is now an actual observation, no longer an inference – but also very nearly closes the inference about why it ended up there. If *Scorpion* is where a simulation based on the cook-off of a torpedo battery puts it, that becomes (see Chapter 3) quite strong evidentiary support for this particular causal theory. The Sontag–Drew chapter concludes with an investigation into all the other evidence for the bad-battery theory, including confidential Navy submarine corps memos, interviews with surviving submarine commanders, and finally (in data declassified only in 1998) pictures of the wreckage showing that the torpedo room compartment had not imploded (meaning that it was already balanced against hydrostatic pressure, i.e., flooded to begin with). The theory was actively rejected by the majority of Navy bureaucrats, because it implied that they themselves were responsible for the loss of the submarine, but this topic takes us too far from the theme of this book.

Let us review. We began with a one-bit fact – *Scorpion* was missing. Agreement (to be called "consilience" in Chapter 2) among the three channels of hydrophone data, as referred to the map of the planned trajectory, led to a location and then a new one-bit fact (reversal of direction) that, via multiple simulations, led to a presumed cause of the catastrophe accompanied by estimates of crucial continuous parameters of the wreckage's glide path. The simulations were confirmed by actual images of the wreckage, showing not only that *Scorpion* lay where the final theory would place her but that her physical form was consistent with one particular version of the detailed explanation of the wreck (explosion in the torpedo room). This is surely as far as you can expect one method (numerical inference) to get you in the reconstruction of historically unique events.

If this ebb and flow of argument seems familiar, it may be because the same cycling and recycling is embedded deep in popular culture as the method of detective stories. Technically obsessed police procedurals by authors such as Ed McBain or