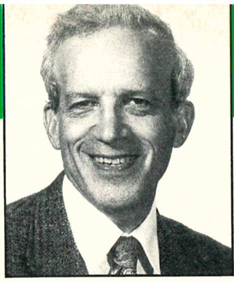
REFERENCE FRAME



FRETTING ABOUT STATISTICS

Daniel Kleppner

A test of whether an anti-apple falls up or down should give a yes–no result, but answers to most experiments are more subtle. Often they lie hidden in stacks of numerical data and come to light only after a grueling search using tools of statistical analysis. A colleague—a first-rate experimenter who probably knows better—claims that if you need to rely on statistics to understand your experiment, you are in serious trouble. The claim is obviously exaggerated, but he has a point: If you need to rely on statistics, you need to worry.

The problem, as every experimenter knows, is that statistical analysis is founded on the assumption of random processes, whereas most experiments are plagued by nonrandom processes. The uncertainties, usually called "errors" (one of the worst misnomers in physics), are often classified as either random errors or systematic errors. In principle, random errors will average out in time. Systematic errors, in contrast, do not go away and may actually get worse. They arise from causes that are probably present but which you cannot control. Distinguishing between systematic and random errors provides the illusion that one knows where the limits of accuracy lie. In practice, the possibilities for systematic errors, or, to put it bluntly, mistakes, are boundless.

Statistical analysis is hardly the special domain of the physical sciences. Economics, epidemiology, paleontology and anthropology are but a few of the many disciplines that depend upon it. Physicists, however, use statistics in a particularly expert fashion, for no other science approaches physics in quantitative accuracy. The loftiest application of statistical thought in physics is the discipline of statistical mechanics, a subject that actually bears little relation to its progenitor in spite of the similarity of their names. The word "statistics," in fact, originated in the lowly context of human affairs rather than from science. It is derived from "statist," a now obsolete term for an expert at statesmanship. Statistics were the numerical facts that statists used to understand the workings of states. The word still carries that connotation, for to the public statistics are facts periodically reported by newspapers to inform us and occasionally reported by politicians to mislead us.

The power of statistics to deceive is so well known that the title "statistician" is slightly suspect. Possibly that is why W. H. Auden, who once proposed ten commandments for professors, included "Thou shalt not sit with statisticians / Nor commit a social science." (All ten commandments can be found in his poem "Under Which Lyre.") Auden's admonition notwithstanding, most experimenters must play the statistician from time to time.

Having raised some reservations about relying on statistics, it is only fair for me to point out that statistical analysis has been crucial to more than a few dazzling discoveries. The anisotropy in the cosmic background radiation recently reported by the Cosmic Background Explorer team is a case in point. (See PHYSICS TODAY, June, page 17.) The heart of the COBE project is a sensitive and incredibly reliable differential radiometer. Even so, discovering the anisotropy required a heroic effort to average out the intrinsic thermal fluctuations arising from noise in the receiver and to deal with a host of artifacts. The radiometer has three pairs of receivers, each of whose twin antennas has a 7° cone angle. Each receiver can achieve a sensitivity of about 15 mK in 1 second. Averaging the fluctuations for a few hours enhances the sensitivity to about 100 μK. To study the cosmic background, the COBE team mapped the sky into a field of 6000 pixels and observed each pixel for a few hours during the course of a year. Altogether, over 70 million measurements were recorded. The anisotropy—a quadrupole distribution in temperature—has a magnitude of only 13 ± 4 μK, much smaller than the noise in any given pixel.

What took most of the COBE team's effort was simply convincing themselves that the statistics were telling the truth. They spent more than one year combing through the data in a search for possible distortions due to the apparatus—for instance, "spring fever" during the occasional periods when the radiometer drifted slightly into the sunshine. The largest problem turned out to be a slight sensitivity to the Earth's magnetic field of the microwave device that switched between the antenna horns. The team's major worries turned out to come not from the apparatus but from nature—everything from heating by the planets to background glow from our Galaxy. To add to the confusion, various sources of Galactic radiation generated noise patches of about 100 μK scattered here and there about the sky.

Not many experiments are as complicated as the COBE project, and data analysis is usually straightforward, at least in principle. Nevertheless most experimenters get nicked one time or another by some malicious trick of the data, and sometimes seriously wounded. I was nicked early enough in my career to learn to approach the business with some respect.

After spending a few years as a graduate student building my apparatus, the time came to try the experiment. The first goal was to search for a resonance pattern that would reveal itself as a large peak in a squiggly line from a chart recorder. I tinkered and

Daniel Kleppner is the Lester Wolfe Professor of Physics and associate director of the Research Laboratory of Electronics at MIT.

played hour after hour. Suddenly, the peak appeared. I tore off the record and rushed to show it to an older and wiser friend who happened to be around at that late hour. He gazed at the record of my triumph and summarized it in one word: "Noise!" He was quite correct. When I "deselected" the pattern of my glorious peak by gazing at a few more feet of the recorder chart, it became evident that the signal looked no more like a resonance curve than the Man in the Moon looks like a man in the Moon. I had seen what I had wanted to see. Such is the power of noisy data: You can see whatever you wish.

Although "eyeballing" the data as I had done can hardly be called statistical analysis, the fact remains that whenever you have an element of random behavior in your data, you can easily fool yourself. With today's cheap and powerful workstations you can accumulate vast piles of data, analyze them in a jiffy and apply sophisticated statistical tests to reassure yourself that the data are consistent and that all is well. What you have really achieved, however, is the ability to fool yourself in a highly sophisticated manner.

The easiest way to let statistics lead you astray is to selectively reject some of your data. Somewhere from the distant past I hear a teacher's voice explain the best method for weighing something: "Take three measurements, average the closest two, and throw away the third." Follow that philosophy and you are sure to run into disaster.

Consider the situation in which you work for months and months to measure some quantity accurately. To convince yourself that your result is absolutely reliable, and perhaps to reduce the final uncertainty a little, you repeat the entire measurement. Many months later you have a new result. Unfortunately, it disagrees with the first measurement by several times the probable error. Typically, you ponder deeply about possible origins for the discrepancy. It is only human to come up with some explanation, such as "if the line voltage dropped during that heat wave in July and there was a magnetic glitch during the thunderstorm like the one we saw once a few years ago...," that would explain it exactly. But in your heart you are not convinced.

Your choice is to bite the bullet and use both results, accepting the fact that your extra work has actually increased your final uncertainty, or to reject one measurement and keep the other. The law governing such a choice is simple: Given two discrepant measurements, the odds are 10 to 1 that whichever you throw out will be the one you should have kept.

Such dilemmas arise in physics in various guises. Sometimes they are blatant, as when a team is searching for a new particle and has a total of ten points that display an unmistakable resonance provided that one or two of the points are disregarded. Sometimes they are subtle, as in an experiment in which you have thousands of data points whose scatter plot forms a lovely Gaussian curve. If the standard deviation is Δ, then the accepted value for the uncertainty in the mean is $\Delta/\sqrt{N}$, where N is the number of points. Everything looks in order provided you maintain a certain degree of ignorance. If you look closely, however, you may discover that your best estimate of the random scatter of single points is less than Δ. Some unknown effect is causing the spread in your data. In such a case, there is no rationale for averaging the data no matter how beautiful the distribution curve looks. Your final uncertainty should be Δ, not $\Delta/\sqrt{N}$. The moral? Ignorance may be bliss, but it is dangerous.

These unsettling thoughts were inspired by a poem on statistics, surely one of the least poetic of all subjects. It sticks in my mind because the poem not only is technically correct; it also comes to grips with some of the deeper human issues that beset scientists. The author is J. V. Cunningham.

Meditation on Statistical Method

Plato despair!
We prove by norms
How numbers bear
Empiric forms,

How random wrongs
Will average right
If time be long
And error slight;

But in our hearts
Hyperbole
Curves and departs
To infinity.

Error is boundless.
Nor hope nor doubt,
Though both be groundless,
Will average out.

* * *

I thank Ray Weiss for explaining to me the intricacies of the COBE project. "Meditation on Statistical Method" is reproduced from The Collected Poems and Epigrams of J. V. Cunningham *(Swallow Press, Chicago, 1971) by kind permission of Jessie C. Cunningham.* ■