An Introduction to Uncertainty and Measurement using the GUM (Guide to the Expression of Uncertainty in Measurement), Les Kirkup and Bob Frenkel, Cambridge Univ. Press, 2006, pp: 233, ISBN 0521844282 (hc), 0521605792 (pbk); Price: US$80/34.99

For the past few years, I have taught a second-year laboratory course on data acquisition and analysis in which the estimation of uncertainties in measured quantities plays an important role. At first thought, such a subject seems "classical" and unchanging, but, in fact, four recent developments have transformed the subject.

The first development is the amazing improvements in data acquisition devices, computer hardware, and software. The humblest laptop computer now has more than enough computer power to carry out almost any conceivable task in a student laboratory. Moreover, routine scientific packages -- Matlab, Igor, LabVIEW, etc. -- now provide built-in routines that implement all these tasks in built-in code. It is now possible to teach an undergraduate with no programming experience how to digitize a signal and send it to a computer, compute standard quantities such as a power spectrum, and display the results -- tasks that would have taken a 1980s graduate student several weeks -- in under five minutes.

The second development is the triumph of Bayesianism as a rational, complete conceptual foundation for the methods of data analysis and inference. The last half of the 20th century was marked by a heated debate between the "frequentist" and Bayesian schools of statistics. The two differ in their definition of probability, with the former interpreting it "objectively" as the ratio of favourable to total cases and the latter defining it as the degree of (rational) belief in a statement. Over the past two decades, Bayesianism has steadily gained the upper hand, as a result of a string of successful analyses not possible by more conventional means. In particular, various computational burdens that made implementation of Bayesian methods difficult have now been resolved (by some of the advances in computer techniques referred to above).

The third development has been the rise of new fields of science in which statistical fluctuations play a more important role than in previous subjects, a development that puts a priority on the mastery of statistical techniques in data collection and analysis. One prominent example is single-molecule biophysics, where inferences must be -- and have been -- teased out of noisy data and where ensemble information is actually deduced from statistics of repeated observations of individual events.

The fourth development has been the extraordinary advances in manufacturing in increasing both control and reliability. The former has led to such high-precision advances as the 45-nm feature size of the latest integrated-circuit chips and the latter to such movements as the "6-sigma" quality-control practices whose goal is to manufacture items with less than 3 "errors," or defects, in a million (6-sigma standard deviation).

It is this last development that seems to have motivated the new book by Les Kirkup and Bob Frenkel, *An Introduction to Uncertainty and Measurement using the GUM (Guide to the Expression of Uncertainty in Measurement)* (Cambridge Univ. Press, 2006). Kirkup and

Frenkel's book is aimed at first- and second-year university laboratory courses and covers traditional topics -- statistical and systematic errors, mean and standard deviations, least-squares analysis, elementary probability distributions, etc. -- with a few twists. The GUM is an effort to standardize definitions and concepts having to do with accuracy, error, precision, and uncertainty in measurement. The effort has been promulgated by organizations such as the Bureau International des Poids et Mesures and the International Standardisation Organisation (ISO). As the authors rightly state, this international effort at bringing a common language to the discussion of uncertainties is largely unknown in the academic physics community.

What is the GUM? Interested readers may consult a free publication (NIST Technical Note 1297), which may be downloaded from the web (http://physics.nist.gov/cuu/Uncertainty/index.html). The GUM introduces a common vocabulary and technical language. Let $y = f(x_1,x_2,...)$ (the "measurement equation"), where the quantity one is trying to measure (the "measurand") is derived from the "input quantities" x_i. One then classifies uncertainties into Type A, those which may be evaluated by statistical methods, and Type B, those which are evaluated by other means. And so it goes. The result is a language that resembles the usual one for discussing uncertainties but uses a common (and somewhat heavy) notation and language for doing so.

For reasons that I explain below, I mostly was disappointed by this book. While it has some nice features, it takes a deeply old-fashioned approach that fails to take into account most of the developments cited above. But I begin with some of the things I liked:

1. There is a good introduction to the culture and lore of the national measurement laboratories, international standards, and the like. The authors work or have close involvement with the Australian National Measurement Institute and are thus well-connected to this community. At least some of the examples should be interesting to students (e.g., photos and discussion of mass standards).

2. The classification of uncertainties into Types A and B, while problematic for reasons discussed below, does lead to a good discussion of how to evaluate errors that are less "formal" but often important. These can include the use of calibration reports, informal appraisals of the quality and reliability of a particular measuring device (including the type of device, the reputation of the manufacturer, etc.). The book gives useful guidelines for turning vague statements into more useful and quantitative forms.

3. The book presents, in a way that is appropriate to the first-and-second-year level of the book, of some less-common but still useful ideas, such as autocorrelation and Allen variance. Indeed, the entire treatment of correlated inputs in Ch. 7 is a highlight of the book, suggesting that data analysis is more than the application of rules and definitions and where the notion the idea that the experimentalist engages in a dialogue with data, trying different methods back and forth until a good understanding is reached.

4. The discussion of the central-limit theorem in Ch. 8 is also good, although for the length of the discussion, it would have been nice to show a counterexample -- the Cauchy

distribution, mentioned in passing in a footnote, is the classic one. A nice problem where that distribution comes up is cited in Sivia's book (see below): the problem of determining the position of a lighthouse that flashes highly collimated light at random times. Because of the geometry, the positions on shore for observations of flashes are Cauchy distributed and averages of the position fluctuate as much as individual observations.

Unfortunately, what I disliked about the book overshadows the nicer features cited above. My main overall complaint is that the book is deeply old fashioned. The important advances mentioned at the beginning of this review get the barest of mentions and have little impact in the overall presentation. For example, Bayesian developments rate only a footnote. As a result, almost every formula is "pulled out of a hat" -- that is, quoted without derivation. This can only reinforce the impression of students that the subject is obscure, filled with rather arbitrary rules. It is true that a book of this level cannot present as much detail than a book for upper-level laboratories. Still, this does not justify the approach the authors take.

As an example where the lack of motivation becomes a serious drawback, the discussion of the "coverage" interval for a Gaussian distribution is long and confusing, while a Bayesian approach is simple and straightforward. The Student t-distribution arises in the estimation of the sampling distribution for the mean when the standard deviation is not known a priori (and is therefore marginalized, or integrated out). The simplification produced by taking a systematic, first-principles approach is striking. (To carry out the integrals involved to derive the distribution is probably too much effort and detail for a second-year course. Still, the Bayesian formulation allows one to understand the underlying idea. Whether one actually carries out the steps of the derivation or not is less important.)

In one case, the limitations of an old-fashioned approach actually leads the authors to advocate a poor analysis technique. At the end of the book, case studies illustrating the application of the material taught in the book are presented. (Overall, this was a nice chapter, giving good examples of how to do a detailed analysis of uncertainty in various applications.) Probably the most interesting of these studies looks at data from an experiment on crater formation. In the experiment, steel balls are dropped on sand from various heights, and the crater diameter is plotted as a function of the ball's kinetic energy. The expected variation is a power law. The data are shown -- and fit -- on a log-log plot. Now, showing power-law data on a log-log graph is fine: our eyes are good at picking out small deviations from linearity. But fitting the logarithm of the data, rather than the data itself is a mistake. To be clear, the "least-squares model" that the book adopts postulates that a set of measured points (xi,yi) are related by yi = $f(x_i) + s_i$, where sigmai are drawn from a normal distribution of mean zero and standard deviation s_i and $f(x)$ is an arbitrary function. Transforming to the logarithm of the measured value transforms the gaussian distribution of errors into a different distribution and invalidates the assumptions of the analysis, leading to biased estimates of parameter values. Once upon a time, linearizing the fitting function was perhaps justified by expediency, in that the numerical routines for minimizing the chi-square statistic were more complicated for nonlinear fitting functions than for linear fitting functions (the linearity is actually in the parameters of $f(x)$, not the function itself). Today, however, standard analysis programs have very good algorithms (typically Levenberg-Marquardt) implemented and there is no important difference in the practical use of one or the other. (Indeed, pedagogically, I prefer to start with the iterative nonlinear methods -- asking students to estimate reasonable starting parameters is better pedagogically than having the value "pop out" of a formula.)

A final example of the old-fashioned approach is the too-limited use of computers and computer-aided data acquisition. The book makes a passing reference to the simplest of Monte Carlo methods, but not enough to be useful. To give a feel for how out of touch the presentation is, at the end of Ch. 10 is a numerical listing of tables -- five pages of random numbers of various sorts, including a thousand samples from a gaussian distribution. Given that my three-year-old laptop can compute and plot a million numbers drawn from a gaussian distribution in one second, I can only wonder at such a waste of paper.

One final issue that I have with this book is one that is true of other books I know: it is not enough to teach methods of data analysis or evaluation of uncertainties. Books should also teach how good experimental design can reduce uncertainties and allow the uncertainties present to be more easily and accurately characterized. In other words, one needs to integrate the design of an experiment with the design of its data analysis. To give but one example, consider the inference of a parameter by least-squares analysis. As mentioned above, one assumes that the measured quantity obeys the relevant law (the f(x) given above) plus random noise drawn from a Gaussian probability distribution. It is less appreciated that by design one can insure that the statistical noise is indeed Gaussian. Let us imagine that the quantity to be measured has a 1 Hz bandwidth -- that is, the fastest change in the signal is on a time scale of 1 sec. One should then measure it at a bandwidth of say 10 Hz. (The Nyquist criterion requires only a 2 Hz signal but it is much easier to analyze a signal if the measured bandwidth is higher.) Traditionally, one might use a digital multimeter (DMM) that records 6.5 digits at 10 Hz. However, it is better to measure the signal using a data acquisition card that digitizes 16 bits at 1 MHz, a standard level of performance. Let us go through the math to understand why. A 16-bit DAQ usually has an effective bit number of about 14 (4.5 digits). (This is the actual signal-to-noise ratio of the card.) If one digitizes at 1 MHz and averages in blocks of 100 000 to get to 10 Hz, the error in the 10 Hz average is reduced by 300, which increases the dynamic range to nearly 7 digits. The dynamic range thus equals the traditional DMM; however -- and this is the important point -- the Central Limit Theorem guarantees that in almost every practical case, the average value reported by the DAQ will obey Gaussian statistics, while in many cases the signal reported by the DMM will not. (The DMM typically uses a sample-and-hold circuit and does not average over the 0.1 second period of a 10 Hz sample rate.) Thus, while the nominal performance of a DMM is equivalent to that of a typical DAQ card, the latter can produce much better results. Not only does the average value reported obey the statistics assumed in least-squares analysis, one also gets a very good estimate of the standard deviation for that average value. With a DMM, estimates of the statistics are much more painful to obtain (measuring 100 000 values at 10 Hz would require three hours!) and might be done at only one point, in the hope that it is "representative" of all the data. Thus, we see that good design can significantly improve the reliability of an experiment. Such "tricks" need to be taught -- students are unlikely to figure them out on their own.

Thus, my overall impression of this book is that it would make a good reference for an instructor, as a resource for ideas, examples, problems, etc. I see no reason to replace our current text, *An Introduction to Error Analysis*, 2nd ed., by John R. Taylor (University Science Books, Sausalito, CA, 1997). While that book is also old-fashioned in its own way, it is more careful pedagogically and gives better motivations for various analysis techniques. At the third- and fourth-year levels, there are two excellent books that are consistent with the viewpoints expressed here: *Data Analysis: A Bayesian Tutorial*, 2nd ed., D. S. Sivia and J. Skilling (Oxford Univ. Press, 2006) and *Statistical Data Analysis*, G. Cowan (Oxford Univ. Press, 1998). At the more elementary first-and-second-year level , we still need a text that integrates the recent technological and conceptual developments in data analysis.

John Bechhoefer
Simon Fraser University
Burnaby, British Columbia, Canada