Understanding and Modelling Variability: Practitioners' Perspectives

Aloke Phatak
CSIRO Mathematical & Information Sciences
Private Bag No. 5
Wembley, WA 6913, Australia
Aloke.Phatak@csiro.au

Geoff Robinson
CSIRO Mathematical & Information Sciences
Private Bag 10
South Clayton MDC, VIC 3169, Australia
Geoff.Robinson@csiro.au

1. Introduction

"Variability is our business", or so we statisticians like to claim. Though it is possible to describe the ways in which mature applied statisticians think about and then model variability in particular scientific or engineering contexts (Wild & Pfannkuch 1999), it is much more difficult to trace how they come to that understanding. Another way of saying this is to ask, "How do students, with all their textbook knowledge and, perhaps, naïve ideas of how to apply that knowledge to real problems, progress to be mature statisticians?" The glib answer is, of course, experience, but it does not explain *how* that process occurs. We explore this question by examining some instances where statisticians and non-statisticians look at the same problem rather differently. From these examples, we can draw some tentative conclusions about the different ways in which statisticians and non-statisticians think about variability. Moreover, they suggest to us that a useful way of hastening students' understanding of variability is to combine contact with real problems – student experiments or projects, for example – with a coherent conceptual framework that emphasizes accepting, describing, and modelling variability.

2. Thinking About Variability

We use the word 'variability' to describe the situation in which observations or measurements ought to be the same, but are not. To begin exploring how statisticians might view variability in a particular context, let's consider the following example.

Measurements of the iron content of successive batches of sugar produced by a crystallization process will almost certainly be different. Statisticians might treat this variability in iron content measurements *as if* it were generated from a probability distribution and then use this model to predict the iron content of the next batch of sugar. Implicit in doing so, of course, is the assumption that the future will behave like the past, that the crystallization process will continue to function in the same way that is has before. There are, no doubt, elaborate deterministic models that could be used to describe how iron content varies with changes in crystallizer operating conditions and feed characteristics. Such models, arising from a detailed knowledge of the kinetics and thermodynamics of crystallization, represent how a particular quantity or *phenomenon*, iron content, is generated. In the real world, however, we can only observe measured quantities, and the key insight of statistics is to view *measured data* as having been produced by a *data generating mechanism* (Lindsey 1996).

One way of expressing this mechanism is to write

$$(1) \qquad \text{Data} = f(\text{true state of nature, noise})$$

There are many elaborations of this, the simplest being that the true state of nature and noise are additive, and both can have complicated structures. Without getting into the philosophical debate about whether we should consider *all* variability as ultimately explainable, we can think of the noise as reflecting variability that, given the context and current scientific knowledge and economic constraints, we *cannot*, or *choose not to*, explain. Several authors (Gould 2004, Lambert 2000, Wild & Pfannkuch 1999) have pointed out that modelling the noise component well is often as important as modelling the true state of nature. Indeed, for some purposes, we may only require a relatively simple model for nature; for example, a simple linear model with covariates might be sufficient to allow operators to make adjustments over a narrow range to a crystallizer to keep it on target. A reliable description of the noise component is required, as Lambert (2000) writes, "to understand what is likely and what is not. It is also needed to understand how reliable the estimated mean or prediction is."

In what ways do statisticians grapple with variability? Wild and Pfannkuch (1999) outline a four-step procedure – noticing and acknowledging; measuring and modelling; explaining and dealing with; and investigating. Most important of all, these steps are carried out within a specific problem context, where the objective may be to generate new knowledge or to make predictions that will be used to make decisions

At the risk of making an unreasonably broad generalization, it seems to us that many people with other types of training, such as mathematicians and engineers, tend to take a more deterministic view of the world. They have considerable subject-matter knowledge that is essential for interpreting data and advancing our collective knowledge, but the flip side of the coin is that, like many students, "they will come up with . . . causal explanations with little or no prompting" (Wild & Pfannkuch 1999, p. 238), even when causal explanations may not be justified. Hence, one of the principal differences between statisticians and non-statisticians is psychological. Instead of seeing large numbers of particular (deterministic) events, such as those that might alter the iron content of the next batch of sugar, a statistician will treat them collectively as *quantifiable* noise. This psychological difference can lead to a different emphasis when addressing the same scientific or engineering problem, and the next section provides some examples.

3. Some Examples

In this section, we illustrate some of the ways in which statisticians and non-statisticians view or treat variability differently. These examples have been drawn from our consulting experience, and therefore some details have been altered to protect confidentiality.

- **Emphasizing Bias** One of us was involved in working with scientists who were developing a biosensor to measure trace concentrations of analyte. After the sensor had been prepared, it was calibrated by measuring its electrical response to several concentration standards. For each standard, an electrical decay curve was measured, and the parameter of interest that was extracted was the maximum slope of the decay. How did the scientists know that the maximum slope *ought* to be measured? They had developed complicated mechanistic models that told them that the maximum slope was a parameter that characterized the response of the sensor to a given concentration of analyte, and hence it could be used to construct a calibration curve.

  The decay data were rather messy, however, and different curve-fitting methods gave different results. To help remove this bias and standardize their analysis, they sought statistical advice. In any long and complex analytical procedure, standardization of data analysis procedures is essential, but it this instance, it quickly became clear from discussions with scientists and technicians that they knew instinctively that the really important sources of variability lay in the preparation of the biosensor – purity and quality of raw materials, sensor design, operator-to-operator differences, environmental conditions, and many others – and that reducing the variability in the response was contingent upon reducing or eliminating these sources. Yet, they continued to focus on reducing the bias in the curve-fitting procedure.

  We would hazard a guess that many scientists, especially those who are involved in making precise measurements of physical properties, tend to focus on bias, sometimes at the expense of more important sources of variability. Lest we be too harsh on the biosensor scientists, we should add that one of us, in a previous incarnation as a bench scientist, spent considerable effort in making minute corrections to dilatometric measurements of the thermal expansion of a particular crystal – corrections for the expansion of the glass dilatometer, temperature fluctuations, etc. – only to realize later that taking into account lot-to-lot variability was far more important than minute bias corrections!

- **Ignoring Variability I** The design and commissioning of mineral processing plants is the preserve of metallurgical engineers, but statisticians, too, ought to become involved. In a recent dispute in which one of us was involved as an expert witness, a mining company sued a consulting engineering firm regarding a mineral processing plant that didn't work as intended. One important issue was that insufficient attention had been given to likely patterns of variability in the input stream. The plant had been designed primarily for the nominal or design grade of input ore, but more attention should have been given to short-term and long-term trends in the many aspects of average ore grade which affect what happens in the plant. This example illustrates the importance of considering variability when a process is on the drawing board, in particular, quantifying the *expected* amount of variability. Though there is uncertainty about the amount of variability, allowing for it might have led to a more robust process design or to the use of blending stockpiles upstream of the processing plant.

- **Ignoring Variability II** One of our colleagues is currently working with endocrinologists on identifying biomarkers for a particular disease. Levels of biomarkers within individuals are predictive of an increased risk for certain diseases; for any individual within a cohort, the risk depends on the decile to which that individual is allocated. For example, the upper quartile may be associated with one disease, the lower quartile with another.

  He recently came across a paper in the endocrinology literature in which the authors expressed surprise that repeat measurements of the level of a particular biomarker for the same individual were quite different. In practical terms, this meant that on the basis of the first measurement, an individual might be assigned to a high risk group; on the basis of the second, to another group altogether. The authors of the paper concluded that variability ought to be taken into account when designing biomarker studies, but they did not (or perhaps were not able to) specify how.

  By partitioning biomarker variability into between- and within-individual variability and then estimating these two quantities, our colleague was able to show that the paper's results were not at all surprising. Indeed, given the estimates of the two components of variability, they were to be expected. In this instance, the scientists working with the subjects and making biomarker measurements understood that somehow, it was important to consider variablity, but they were not able to articulate that qualitative understanding into statistical terms. By contrast, an experienced statistician was immediately able to grasp the essential elements of the problem and draw sensible conclusions about the expected variability between repeat measurements.

- **Differences Between Study and Target Populations** One of us was involved in the problem of finding the relationship between the readings of a set of strain gauges and the weight of a vehicle which travels on a road pavement into which the strain gauges have been inserted. Someone with a deterministic view might be concerned only with finding the equation which best describes the relationship between the readings of the strain gauges and the vehicle weight. An experienced statistician will also be interested in the precision of this relationship and the robustness of this precision: whether the precision might be different for other road pavements, or might change according to the way the strain gauges were inserted into the pavement. He/she should also be interested in considering whether there is enough information available to make a useful prediction about the likely future precision of such devices or whether the amount of uncertainty is such that more data should be collected. If more data is to be collected then the expenditure of such data-collection effort should be guided by consideration of which components of uncertainty most need to be reduced.

These and other examples suggest to us that scientists and engineers, those who are best placed to *observe* variability and who know viscerally that it exists, are often the ones in whom two apparently irreconcilable ideas exist simultaneously. On the one hand, their scientific training tells them that *if only* they knew the (deterministic) model exactly and could measure its inputs precisely, they could predict outputs exactly; on the other, their contact with the messiness of real processes and data tells them that 'perfect' knowledge is neither attainable nor necessary. These two ideas could be reconciled by a better understanding of variability: acknowledging that it exists; being able to model it; and then drawing justifiable conclusions from the analysis. To non-statisticians, that realization can be a revelatory one.

4. Helping Students Become Good Statistical Practitioners

We remarked much earlier that 'experience' is the answer that first comes to mind when we think about our own transition to mature statisticians. But what is it about experience that hastens the process? First of all, context. The problems we tackle as practicing statisticians are not textbook ones, but ones in which we assist in generating new knowledge or in making decisions that have consequences for people or enterprises. As a result, statisticians have to think carefully about *justifying* their conclusions, and so are forced to think about variability. Second, we also want our conclusions to be generalizable, and so in designing an experiment, for example, we may include design factors that we expect will vary in the future.

Context is also important to help students think about variability. We agree with Wild and Pfannkuch (1999, p. 224), however, that it is not enough to "let [students] do projects," and the examples in the previous section demonstrate that even individuals with considerable subject-matter knowledge who are working on real problems often misunderstand the importance of considering variability. We also believe that contact with real problems – students' own experiments and projects, for example – has to be accompanied by a coherent conceptual framework. Some aspects of that framework include:

- understanding that describing phenomena in terms of a few components of variability can be simpler than describing them in terms of large numbers of deterministic events;
- describing and modelling variability as well as the mean; and
- eliciting information about what people don't know and estimating how variable process outcomes are likely to be.

Additionally, there are strategies for putting tertiary statistics students into situations where they can serve a 'statistical apprenticeship' and learn by working with more experienced statisticians. Our organization, for example, hires third-year students over the summer vacation, and they work either on consulting problems or on research arising from consulting. Other strategies include innovative courses on statistical consulting (Taplin 2003), and co-operative or 'sandwich' programs where statistics students work in organizations or enterprises.

Paradoxically, it may be easier to teach science and engineering students about variability because such students are always doing experiments. Though it is true that many of these experiments are demonstrations of well-established concepts rather than open-ended investigations, it is still possible to incorporate notions of randomization, replication, and blocking in simple experiments where the objective may be to calculate the physical properties of a substance or material. For an attempt to do so, see Burke, Phatak, Reilly, and Hudgins (1993).

In discussions with our colleagues, they have all remarked that it is hard to turn the clock back and put themselves in the frame of mind in which they did not think 'naturally' about variability. In that sense, their (and our) past and current understanding of variability are a bit like 'incommensurable paradigms' (Kuhn 1970). Consequently, it is difficult to trace the path from a naïve to a mature conception of variability. By examining how some non-statisticians and statisticians think about variability, we have tried to examine two points along that path, but it is clear that there is considerable work yet to be done.

Acknowledgments

We have discussed many of these ideas and traded many stories with our colleagues within and outside CSIRO. In particular, thanks go to Ian Saunders, Ross Sparks, Doug Shaw, and Jodie Thompson.

REFERENCES

Burke, A., Phatak, A., Reilly, P., & Hudgins, R. (1993). Introducing statistical concepts in the undergraduate laboratory: Linking theory and practice. *Chem. Eng. Educ.*, **27** (2), 130–135.

Gould, R. (2004). Variability. One statistician's view. *Stat. Educ. Res. J.*, **3** (2), 7–16.

Kuhn, T. (1970). *The Structure of Scientific Revolutions*. 2nd edition, Chicago: University of Chicago Press.

Lambert, D. (2000). What use is statistics for massive data? (http://cm.bell-labs.com/stat/doc/statistics.ps).

Lindsey, J. (1996). *Parametric Statistical Inference*. Oxford: Oxford University Press.

Taplin, R. (2003). Teaching statistical consulting before statistical methodology. *Aust. J. Stat.*, **45** (2), 141–152.

Wild, C. & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *Int. Stat. Rev*, **67** (3), 223–265.

RÉSUMÉ

Contrairement au point de vue déterministe des scientifiques et des ingénieurs, la pensée statistique souligne l'importance de la variabilité. Dans cet article, nous présentons plusieurs cas d'études pris de nos dossiers de consultation statistique, et à l'aide de ces exemples nous tirons quelques conclusions au sujet des manières différentes de penser à la variabilité. De plus, ces différences nous suggèrent une façon d'accélérer la compréhension de la variabilité auprès des étudiants .