In Theory...

Assessing 10-50-90s: A Surprise

Bob Clemen

Many decision analyses require eliciting from experts probability distributions for continuous quantities. A reasonable starting point is to ask an expert to give 10th, 50th, and 90th percentiles (so-called 10-50-90s). We know from the behavioral decision theory literature, however, that people tend to be overconfident; their 10-90 ranges tend to be too narrow. Although experts appear to be slightly less overconfident, the effect is pervasive. Lichtenstein, Fischhoff, and Phillips (1982) provide an extensive overview of the early research documenting the overconfidence effect. For other reviews, see von Winterfeldt and Edwards (1986) and Yates (1990).

Overconfidence is thought to derive at least in part from an anchor-and-adjustment heuristic. The heuristic is that the expert chooses a "best estimate" and then adjusts up or down from that estimate in order to get the extreme percentiles. The adjustment is typically inadequate, leading to intervals that are too narrow. From this argument derives some typical admonitions. First is to have the expert think hard about what kinds of scenarios could lead to different results, especially extreme results (Spetzler and Stael Von Holstein 1975). In a similar vein, Morgan and Henrion (1990) and others argue for a "conditioning" stage in a probability-assessment protocol, during which an expert organizes and expresses what he or she knows about the problem at hand.

A specific procedural admonition is to have the expert think about the extremes first, assessing the 10th and 90th percentiles first, and then the 50th. Doing so, the argument goes, should reduce the anchoring effect and thereby result in better calibrated 10th and 90th percentiles (Clemen 1996, p. 284). Russo and Schoemaker (1992) report the results of an experiment that substantiates this argument.

In a recent decision analysis class, I thought I would demonstrate this effect to my students. So I had them respond to a fairly typical "overconfidence quiz" with several almanac questions. Half of the students provided 10-50-90s, assessing the median first, and half assessed only the 10-90s. To my surprise, the 10-90s showed worse calibration; that is, the actual values fell inside the assessed interval less (49% versus 57% of the time).

Still believing that the anchor-and-adjust story was correct, I went to the literature to find experimental evidence documenting what I believed to be true. Guided by suggestions from my behavioral decision theory colleagues, I was in for even more of a surprise. The earliest experimental evidence I could find comes from Judith Selvidge's dissertation (Selvidge 1975, 1980). She found that assessing "central" values first (median and quartiles) and then the extreme values (1st and 99th percentiles) gave better calibration than assessing the extreme values first. (In fact, her results showed an improvement in the hit rate of about 8 percentage points, exactly the same as my classroom results.) Selvidge argued that "naming the extremes immediately causes an assessor to fix on values that roughly define the range of the UQ but that correspond to fractiles less extreme than those requested" (Selvidge 1980, p 500).

More recently, Block and Harper (1991) found the same result. Juslin, Wennerholm, and Olsson (1999) go further and explain this result by a "priming" effect. That is, assessing a point estimate first means that the expert performs more information processing about the uncertain quantity before assessing the probability interval. This explanation extends Selvidge's by including a cognitive mechanism for the effect.

It appears that my classroom results, although a surprise to me, were not inconsistent with existing experimental results. But what about the other side? Are there any studies showing the result I originally expected? Russo and Schoemaker's (1992) study is the only one that I found. Although I do not doubt their results, they may be in the minority.

What conclusions can we draw from this exercise? First, the typical DA admonition to assess extremes first and best estimates last may be bad advice. (See Fred Rolle's contribution on this topic in this newsletter, though. We do not know for sure what works best with knowledgeable experts!) Second, although that admonition is based on the well-accepted anchor-

continued on page 15

... In Practice

Assessing Probabilities at DuPont – Maybe Not A Surprise

Fred Rolle
DuPont

How are we doing at DuPont when it comes to talking about uncertainties in the context of making complex decisions? Good progress! The language of 10-50-90 is well known as a result of our DA courses and the fact that we have been practicing DA here for more than 10 years. Early on, there was occasional resistance to owning up to uncertainty, but that has diminished with experience and management's increasing interest to understand risks associated with pending decisions. We are going to be assisted by the aggressive Six-Sigma program that is being implemented across the company. Variation and uncertainty are an integral part of Six-Sigma, and people are becoming more used to discussing such topics. This will help accelerate the embracing of uncertainty when complex business decisions are being made.

Uncertainties faced in our typical decisions involve areas in which the business team has familiarity and experience. The subject matter experts are employees of the business, who are looked-to for expertise in judging the future. When assessing uncertainties from our experts, we normally encounter these issues:

Expert bias – *the need to be relatively certain.*

Anchoring – s*ingle-point forecasts have previously been made in the business for a variety of reasons, sometimes with an optimistic tilt.*

Both result in the situation described by Bob Clemen: 10-90 ranges that are too narrow.

Assessing the 50% value first is helpful when the expert needs to get "calibrated", but this is rarely the case when discussing uncertainties that the expert knows well. It appears that Bob Clemen's research with his class and the literature he cited dealt with uncertainties not familiar to the expert. In those cases, the 50% assessment may provide an essential grounding, thus enhancing ability to then judge degree of uncertainty. But, that is not typical of what we face. We have found that the most effective approach is to de-anchor, by discussing extreme outcomes with the expert and then developing the 10% and 90% outcomes before doing the 50% value.

Experts need help in thinking about the 10%/90% values. To assist, we ask the expert to choose between a game with a 10% of winning and a game that pays off if the outcome in question would turn out to exceed some value chosen by the interviewer. This value is then changed a few times until the expert is indifferent between the two games. After this approach is used once or twice, the expert usually can think directly about the 10%/90% values, and asking the expert to choose between the two games is discontinued for additional uncertainties.

If the 50% value were assessed first, it is very likely the value would be what had been forecast or assumed for planning purposes at some earlier time by the business. Doing the 10%/90% values first has the double benefit of sharpening the thinking on the important median value. This sequencing is favored when the expert really is an expert. When the "expert" really is not an expert, the reverse sequencing (median value first) would make sense.

General DA question? Send it out to the DAS by emailing it to *DAList@fuqua.duke.edu*

Interesting spin on a DA issue?

Pair off with a friend and try your hand at *"In Theory/In Practice"*

Help Wanted

Continued from page 7

Decisions (1995) has a specific business school orientation and focuses more on formal modeling. Keeney's Value-Focused Thinking (1992) is primarily a discourse on decision aiding principles. Watson and Buede's Decision Synthesis (1987) gives a deeper treatment of decision analysis principles. Except, perhaps, for Behn and Vaupel, the texts are more complementary than competitive and can be assigned as supplementary reading. This text is distinguished from all five books in that: it anchors technique to decision aiding practice (with "how-to" guidance); it focuses on the intended use and user (rather than only on the decision problem itself); and it includes "hands-on" projects as homework assignments.

There are no course prerequisites and no explicit mathematics is used, however, some quantitative aptitude or experience is desirable. I have been teaching both parts A and B in a single course to grad students for several years, in psychology at LSE and public policy at George Mason. Presently, I am looking for other people to appraise the book and give me feedback before it goes final. Please contact me at *rbrown@gmu.edu* if you have any interest. I will soon be putting it on the DAS website, but in the meantime I'm happy to feed people directly.

Don't forget to make your contribution(s) for the next issue of DASN!

In Theory...

Continued from page 2

and-adjust heuristic, it appears that this explanation may be too naïve. The subjective judgment process is complicated, and we should be careful not to oversimplify.

How will I proceed when directing future elicitations? I will certainly continue to have the expert think hard about possible scenarios, especially those that could lead to extreme results. And after the assessment, I will ask him or her to reconsider the extremes. But for now, I intend to have the expert give a best estimate first and follow up with the 10-90s.

References

Block, R. A., & Harper, D. R. (1991). Overconfidence in estimation: Testing the anchoring and adjustment hypothesis. *Organizational Behavior and Human Decision Processes, 49,* 188-207.

Clemen, R. T. (1996). *Making hard decisions, 2nd Ed.* Pacific Grove, CA: Duxbury Press.

Juslin, P., Wennerholm, P., & Olsson, H. (1999). Format dependence in subjective probability calibration. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1038-1052.

Lichtenstein, S., Fischhoff, B., and Phillips, L. D. (1982). Calibration of subjective probabilities: The state of the art up to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306-334). New York: Cambridge University Press.

Morgan, M. G., and Henrion, M. (1990). *Uncertainty*. New York: Cambridge University Press.

Russo, J. E., and Schoemaker, P, J. H. (1992). Managing overconfidence. *Sloan Management Review, 33,* 7-17.

Selvidge, J. E. (1975). Experimental comparison of different methods for assessing the extremes of probability distributions by the fractile method. *Management Science Report Series*, University of Colorado, Graduate School of Business, Business Research Division.

Selvidge, J. E. (1980). Assessing the extremes of probability distributions by the fractile method. *Decision Sciences, 11*, 493-502.

Spetzler, C., & Stael Von Holstein, C.-A. S. (1975). Probability encoding in decision analysis. *Management Science, 22*, 340-358.

Von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research*. New York: Cambridge University Press.

Yates, J. F. (1990). *Judgment and decision making*. Englewood Cliffs, NJ: Prentice-Hall.

Economics and Finance

Continued from page 14

Robert W. Grubbstrom (Linkiping Institute of Technology), Steven Haberman (City U. of London), Jean-Yves Jaffray (U. Paris VI), Panos M. Pardalos (U. of Florida), Eugenio Regazzini (U. of Pavia), Wolfgang J. Runggaldier (U. of Padua), Aldo Rustichini (Boston U.). Submissions can be sent to the Editor: Marco LiCalzi, Dept. of Applied Mathematics, U. of Venice, Dorsoduro 3825/e, 30123 Venice, Italy. More info at http://link.springer.de/link/service/journals/10203/