Numeracy Nugget #6: Modeling the Future - For Whom the Bell Curve Tolls

George Rebane

More and more public policy is based on poorly understood computer models of very complex processes like regional air quality, local traffic congestion, and worldwide climate change. Paul Sieving wrote an excellent yet very brief article (The Union's *Other Voices*, 6 March 2007) that highlighted how the output of such predictive models may lead us to wrong conclusions and, therefore, bad policy. However, it turns out the picture he painted is too rosy, this perhaps because he tried to keep the message short – a sad requirement of the times.

Exposing the next important layer of complex systems modeling forces us to understand the difference between deterministic analytical (one-shot) models and stochastic iterative models. For illustration, a simple deterministic analytical model of temperature as affected by CO2 levels is T = 51.7 + 1.2*C where C is some predicted concentration of atmospheric CO2. Just plug in a value of C and you get the predicted temperature, no problem. Only there are many problems since we don't know whether 51.7 and 1.2 are really correct because past data shows these 'constants' to vary from 47.9 to 53.0 and 0.8 to 1.7. And, of course, we're really not sure that there's not more to the model than a simple multiply and add, let alone what errors we may encounter in selecting a value for C. So you can see that the resulting temperature predictions may be all over the place even with such a simple model.

System scientists try to overcome a part of this problem by first making their models iterative. This means that instead of predicting fifty years into the future in one shot, they try to do a better job at just predicting what the temperature will be, say, tomorrow. An example of a very simple iterative model then is T(tomorrow) = 0.92*T(today) + 0.08*[C(tomorrow) – 23.4]. So if we feel real confident about predicting the future levels of CO2 over the next fifty years, we can simply measure the temperature today, calculate tomorrow's temperature, use that to calculate the temperature for the day after tomorrow, and so on for fifty years worth of days. Here all the same concerns about the numbers used and actual form of the model are still with us, since if we change anything by even the slightest amount, we will get wildly different results at the end of the fifty years.

In its simplest form a stochastic process is one composed of lower-level random processes. For example rolling one die is a random process giving a number from one to six each with equal likelihood. Then adding the number on the die to the number from the spin of a roulette wheel is the output of a stochastic process. Actually, realworld stochastic processes contain not only simple random processes but also other stochastic processes all nested like the brightly painted Russian matrioshka dolls.

In our temperature example above, for each day's iteration we must now draw a number that multiplies T(today) from a distribution (Bell curve) of numbers whose mean may or not be 0.92, and the same for the other numbers (more Bell curves) in the iterative model. Also, C(tomorrow) will most likely be the output of another stochastic iterative model (even more Bell curves) that may even use predicted temperatures from our model as an input thereby making these models coupled. Actual climate models have hundreds of such equations that are coupled over climate variables and geographical regions.

You can now see how difficult is any attempt to reliably model the future. Because of such things and more, science teaches that most of the future is inaccessible to us. Nevertheless, politicians, ideologues, and demagogues all claim to save the layman the trouble of these inconvenient details by simply quoting one number that best serves their purposes – 'Scientists tell us that we are headed for temperatures 10 degrees higher in another twenty years'. This number, wrapped in such a meaningless statement, is always obtained by playing loose and fast with the model *du jour* assuming that the model is even exercised. A technically correct and socially useful result is never quoted as a single number but as a stated range of values with the attendant probability of the actual value falling into this range - 'The 90% range of predicted average temperature is from 68 to 82 degrees Fahrenheit for the year 2030.' Such a report should also have available (perhaps on a website) the model used along with the published standard input dataset, and any special assumptions.

Unfortunately, all of this is deemed indigestible for the public at large who must assess the truth of any proposition from the more popular indicators such as the emotional level and repetition frequency of its presentation. This is how we guide the ship of state.

… and those who never attempt to know
for whom the bell curve tolls;
it tolls for thee.

Solution to NN5 Problem: Problem restated: Three contestants are blindfolded. A white piece of paper is pasted on each contestant's forehead. They are told 1) that they each have a white or black paper on their forehead, 2) that not all the pasted pieces of paper are black, and 3) that the first one to correctly deduce the color of their paper will win the prize. The blindfolds are removed simultaneously and after a second or so all three contestants announce 'White!' at the same time. Why?

The contestants A, B, and C can each reason as follows: A thinks 'Suppose my paper is Black, and since B and C have White, therefore B would say 'If I and A both have Black, then C would immediately announce White, but C hesitates, therefore I (B) will announce White', but B is also silent, therefore I (A) must not have black and I will announce White.' All three take the same time to reason identically and then call out 'White!'. (This is Problem #278 of the 1972 classic *Moscow Puzzles* collection by B.A. Kordemsky edited by Martin Gardner. This important collection of brain recreations entertained two generations of Soviet youngsters and helped produce the technicians who contributed to the cold war between the USSR and the West. They are now in worldwide pool of technologists of the Russian Federation who compete with the dwindling supply of American technical graduates.)

An important point to remember in attacking problems of uncertainty is to just assume that an unknown proposition has a certain value or condition, and see if the other known propositions are then consistent with that assumption. If this doesn't resolve the problem, then continue 'supposing' about each of the other unknown propositions (singly or in groups) until you start making progress by eliminating some of the uncertainty about the unknown proposition you are tackling. Proceed in this manner until things (hopefully) sort themselves out. Solving the popular puzzle Sudoku is a good example of this process. Of course, all realworld problems may not yield to this or any kind of reasoning because there is not enough information available from the start or the number of plausible solutions (the 'search space') is too large. But the above process is still useful because you can identify what additional information to seek by reasoning 'if I only knew that, then this would have to be so'.

NN6 Problem: You pay $1,000 to play a game of dice consisting of a maximum of twenty throws. The game ends when you throw double sixes ('boxcars'). Each time you miss throwing boxcars you lose $50 from your original $1,000 investment. When you hit boxcars the casino will double the money you have left. For example, if you threw boxcars on the third roll you would have $900 added to the $900 you had left after the first two rolls and walk away with a total of $1,800. What are your odds of walking away a winner? What is the amount you would expect to win in this game (i.e. the average winning after playing the game a large number of times)?