Four types of probabilities

By David B. South

The term "probability" has several definitions. I prefer the following: a quantitative description of the likely occurrence of a particular event. Probability is often expressed on a scale from 0 to 100% but researchers often use a scale of 0 to 1; a rare event has a probability close to 0 while a very common event has a probability close to 1. Probabilities are used as a tool to support conclusions regarding both controlled and "natural" experiments.

I deal mostly with probabilities derived from controlled and natural experiments but only recently did I realize there are other types of probabilities. My reason for writing this essay is to point out that not all probabilities are created equal; some are simply guesses. Unfortunately, information about the origin of the probabilities has not been communicated clearly or consistently (NOAA 2009).

In my opinion, probabilities obtained quantitatively are more meaningful than those obtained by guessing or those just based on logic. In one case, I thought the value was obtained by using a formula, but it took me months to realize that the terms "very unlikely" and "likely" were simply guesses. Some argue that "it is insufficient to describe uncertainty in terms of qualitative language, using words such as 'likely' or 'unlikely'"(NOAA 2009). Unfortunately, subjective guesses (which are not qualitative) are sometimes presented as probabilities by researchers who advocate for certain policy changes and this has the potential to confuse policymakers. For this reason, I suggest some clarification of probability terminology is needed.

Type A: Probabilities derived from Controlled Experiments

A controlled experiment generally involves comparing a treated sample with a control sample (which is practically identical to the treated sample except for one variable). Consider a nursery study where herbicide treatments are replicated five times. Visual observations suggest the treatment stunted seedlings. To determine if the herbicide treatment caused a reduction in biomass, I would collect biomass data and then calculate the probability that the observed difference (between treated and untreated seedlings) could occur by random chance. I will have confidence in my conclusion (ie. the herbicide stunted seedlings) if the probability of me being wrong is less than one in 20 ($p<0.05$).

Type B: Probabilities derived from Natural Events

In some cases, researchers collect data from observations made on the natural systems. For example, when tip-dieback is observed in young pine plantations, is it associated with high levels of potassium in the terminal? To address this question, foliage samples of terminals could be taken from 20 trees that had tip-dieback and these could be compared to samples from 20 trees than had no dieback. A t-test could be conducted on these data. If the probability value is less than 0.05, this would be sufficient to reject the null hypothesis (Ho: the observed difference in potassium concentrations is due to chance alone). However, results from a survey like this say nothing about cause and effect. Just because tip-dieback might be associated with high potassium, this does not mean that high potassium levels are the primary cause of tip-dieback. Surveys of this type are often prone to confounding.

Type C: Probabilities derived from "Complex Computer Models"

Many natural systems are very complex. Therefore, controlled experiments are sometimes prohibitively difficult, or impossible to carry out, or require much time. For example, testing the long-term effect of weed control treatments on pine growth may require two decades of data collection (e.g. Lindsay et al. 2009). Therefore, it is much easier and quicker to simulate the real world using computer models.

A few researchers calculate probability values using complex computer models. Some will run an analysis "over and over again on a fast computer, using different input values, from which it is possible to compile the results into probability distributions." This approach is termed "stochastic simulation" (NOAA 2009). If 1,000 simulations are conducted, and X occurs 5% of the time, then the TYPE C probability of X is reported as 5% (note: this is not the same as a p-value). I should point out that in cases where complex models contain many parameters, accurate predictions using the model become problematic.

However, in a few rare cases, I have seen researchers calculate a p-value to determine if input variables in a complex computer model caused a significant difference in the output (i.e. no real data were involved). For example, I created the following example from examining output from a loblolly pine growth and yield program. In this case, I wanted to know: would two years of herbaceous weed control cause a statistically significant increase in green tons of pine logs at age 20 years? The computer model was run ten times (all with no hardwood competition). Five scenarios were conducted with herbaceous weed control and five scenarios run with herbaceous weeds. Surprisingly, on average, the scenarios with herbaceous competition produced about 4.6% more biomass at age 20 years than scenarios with 2 years of weed control. A t-test indicated a probability of 0.2683 (or about 1 in 4 chance of the difference occurring by chance). Although this TYPE C probability was not significant ($\alpha = 0.05$), it does not prove that herbaceous weed control does not increase yields in real plantations. In fact, from 11 sites across the South (with no hardwood competition), the presence of herbaceous weeds decreased biomass in all 11 cases (South et al. 2006). This illustrates that conclusions obtained from a TYPE C probability could differ dramatically from those obtained from a TYPE A probability. This is because sampling done in the virtual world of complex computer model is not the same as sampling conducted in the real world.

Many researchers who work with stochastic simulations do not report p-values because when generating numbers with complex models, the p-value is simply a function of the number of simulations. In many cases, statistically significant p-values can be generated simply by running a greater number of simulations. This is especially true when the researcher already knows the variable in question is an integral part of the model.

Type O: Probabilities derived from a "Show of Hands"

When researchers do not agree about whether X or Y is correct, then policymakers may want some clarity as to what policy is best when predicting an uncertain future. In some cases where uncertainty levels are high, policy makers will ask a number of leading experts to provide their "judgments" in the form of subjective probability distributions (NOAA 2009). This process is known as "expert elicitation." Likelihood can be used to characterize the probability of a future event and may be based on an elicitation of expert views. For example, if most volcano experts think the probability of an eruption is greater than 90 percent (for the next year), they might say it is "very likely" and people should move to avoid the danger. In some cases, policy makers might interview a number of experts and then use different methods to determine which opinions should be followed (e.g. move or not move) (Winkler 1968). However, regardless the method used to choose which researchers get to set policy, all TYPE O probabilities are subjectively determined.

Policymakers should not be mislead

When researchers become policy advocates (Nelson and Vucetich 2009), they should not mislead policymakers into thinking all probability values carry the same weight. Some probabilities represent actual events from the past (e.g. TYPE A and TYPE B above) while some involve predictions made using output from computer models (Type C) or opinions (TYPE O). Researchers should make it clear when probability values are based just on opinions, on output from computer models, or involve observations from the real world.

For example, let us consider the following quote from "The Independent" (27 June, 2008): "Polar scientists reveal dramatic new evidence of climate change." It turns out this "new evidence" was simply a prediction that a record low extent of ice would occur in 2008. This prediction (or "new evidence") was supported by a probability value of 59%. Was this value derived using stochastic simulations (TYPE C), or was it a guess made by polling researchers (TYPE O), or was it based on an extrapolation of a simple regression line of observed satellite records? Although some arctic researchers predicted a record low of 3.1 million sq. km of Arctic ice in 2008 (Rigor et al. 2008), the value that year did not go below 4.6 million sq. km. This 1.5 million sq. km miss supports the view that many environmental scientists cannot predict the future with precision (Oreske et al. 1994; Pilkey and Pilkey-Jarvis 2007). I wonder how many policymakers thought the "59 percent chance" was approximated simply by dividing the number of previous Arctic sea ice records by the number of years of satellite data? The next time you hear someone make a forecast using a probability value, ask them…. what TYPE of probability are you using? The answer could be very informative.

References

Lindsay, A.; Oester, P; Cole, E. 2009: Twenty-year response of ponderosa pine (Pinus ponderosa Laws.) to treatment with hexazinone in northeastern Oregon. *Western Journal of Applied Forestry* 24:151-156.

National Oceanic and Atmospheric Administration (NOAA). 2009: Best practice approaches for characterizing, communicating and incorporating scientific uncertainty in climate decision making. U.S. Climate Change Science Program. Synthesis and Assessment Product 5.2. 156 p. www.climatescience.gov/Library/sap/sap5-2/sap5-2prospectus-final.pdf

Nelson, M.P.; Vucetich, J.A. 2009: On advocacy by environmental scientist: what, whether, why, and how. *Conservation Biology* 23: (in press).

Oreskes, N.; Shrader-Frechette, L.; Belitz, K. 1994: Verification, validation, and confirmation of numerical models in the earth sciences. *Science* 263: 641-646.

Pilkey, O.H.; Pilkey-Jarvis, L. 2007: *Useless Arithmetic: Why Environmental Scientists Can't Predict the Future*: Columbia University Press, New York, 230 p.

Rigor, I.G.; Clemente-Colon, P; Nghiem, S.V.; Wood, J.; Brinkley, J.; Arbetter, T. 2008: 2008 Outlook for Arctic Sea ice this summer. http://seaice.apl.washington.edu/Outlook/2008/

South, D.B.; Miller, J.H.; Kimberly M.O.; VanderSchaaf, C.L. 2006: Determining productivity gains from herbaceous vegetation management with 'age-shift' calculations. *Forestry* 79:43-56.

Winkler, R.L. 1968: The consensus of subjective probability distributions. *Management Science* 15:B61-B75.