Bringing Environmental Policy and Regulation into the 21st Century, Part 2 (Newsletter)*

November 23, 2015

The null hypothesis/significance testing (NHST or frequentist) analytical paradigm does not produce answers for environmental policy or regulatory decisions because rejecting the null hypothesis (of no difference between data sets) says nothing about why or by how much they differ. The likelihood (information theoretic) paradigm overcomes many of NHST's problems and can be applied to environmental data when its limitations are understood.

The NHST approach tests how well the data fit a single null hypothesis. The likelihood approach tests how well multiple hypotheses fit the data and identifies the hypothesis that maximizes the likelihood of explaining the data, hence its common name of Maximum Likelihood Estimation (MLE). The difference between testing data for fit to an explanation and testing explanations for their fit to data are critical for policy and regulatory decisions. So is understanding what a statistical hypothesis represents.

Data are collected (sampled) to understand the population of which they are a small part because we cannot measure or observe the entire population. Think of a political poll that reports support for a candidate based on telephone conversations with a very few people. Pollsters use a hypothesis – a probability distribution – that allows them to extrapolate from their tiny sample to the entire population and present the confidence limits that the answer falls within a defined value range with a specified degree of confidence. Statistical hypotheses are probability distributions describing the entire population from which we have collected a sample. The NHST paradigm most commonly assumes the population is normally distributed (the familiar "bell-shaped" curve), yet environmental data almost never are normally distributed.

The MLE approach overcomes this limitation by using a likelihood function to test a set of probability distributions (e.g., Log-normal, Binomial, Poisson, Gamma) for how well each describes the data. When the probability distribution (hypothesis) with the maximum likelihood is identified the properties of that distribution are used to make predictions about the entire population,

*Copyright ©2015 Applied Ecosystem Services, Inc.

compare data sets to determine if they come from the same population, make predictions, and inform regulatory and policy decisions.

An example of applying the MLE approach to environmental data is characterizing chemical concentrations when some cannot be measured because they are below the chemical laboratory's method detection limits. Since these values cannot be quantified (they are unknown) any arbitrary value assigned to them is wrong. One appropriate statistical model for these data fits the measured concentrations to several probability distribution curves (hypotheses) and identifies the one with the maximum likelihood of describing those data. The unknown data are then spread along the left side of that curve, below the detection limit. Using the mathematical characteristics (center, spread, skewness) of that distribution allows the data to be described, characterized, and compared with other data to look for change over time or space.

Applying the MLE approach to environmental data allows policy makers, regulators, and the regulated public to make better informed decisions that are technically sound and legally defensible. However, there is still more information that has value when making these decisions: prior knowledge.

Unless data are from a site that has not has not been examined before, there is a history of data from that location. Prior knowledge is valuable in decision making in the business and financial worlds (think credit scores when applying for a loan) and is equally valuable when examining environmental data for policy and regulatory decisions. The statistical paradigm that includes prior knowledge is based on the Rev. Thomas Bayes' theory of conditional probability. Thus, it is called the Bayesian paradigm.

–

Rich Shepard / Applied Ecosystem Services, Inc.
503-667-4517 / www [dot] appl-ecosys [dot] com