

Comment: Fuzzy and Bayesian p -Values and u -Values

Andrew Gelman

It is a pleasure to discuss this fascinating paper, which presents a new way to express uncertainty in p -values for composite null hypotheses. I particularly like the graphical display above equation (1.2). I have some brief comments about the confidence interval for binomial proportions and then some more to say about the relationship to Bayesian p -values.

Geyer and Meeden's fuzzy p -values are related to Bayesian p -values in that they recognize the uncertainty inherent in testing a composite null hypothesis. Fuzzy and Bayesian p -values also both allow for test statistics to depend on missing data, latent data and parameters, as well as on observed data. There is the potential to generalize these ideas to graphical test statistics and connect to exploratory data analysis.

INFERENCE FOR BINOMIAL PROBABILITIES

For the practical problem of inference for binomial probabilities, I believe methods such as described by Agresti and Min (2005) should work well. I could imagine standard practice being to use such an approximate method, supplemented by a fuzzy interval, to give a sense of the range of possibilities. In more complicated discrete-data models such as logistic regression, I could once again imagine fuzzy p -values and intervals playing a useful role in conveying the sensitivity of likelihood-based estimates.

FUZZY AND BAYESIAN p -VALUES

I am sympathetic to the authors' desire to define fuzzy p -values non-Bayesianly. There are certainly going to be cases where a Bayesian approach is not desired. From a Bayesian context, a posterior p -value is the probability, given the data, that a future observation is more extreme (as measured by some test variable) than the data. Mathematically, a Bayesian p -value can be computed by averaging over the distribution

Andrew Gelman is Professor, Department of Statistics and Department of Political Science, Columbia University, New York, New York 10027, USA (e-mail: gelman@stat.columbia.edu).

of p -values (with distribution induced by uncertainty about unknown parameters, except in the special case of a pivotal test statistic); see, for example, Gelman, Meng and Stern (1996) and Bayarri and Berger (2000). In contrast, the fuzzy method considers the distribution without averaging over it. This is similar to the Bayesian giving you a single number, but at the cost of requiring a prior distribution. It is good for both sorts of methods to be available.

Fuzzy confidence intervals and p -values are not the same as Bayesian posterior intervals and p -values (except in the special case of pivotal test quantities), but there are similarities, and indeed there could be hybrid versions that average over some parameters and condition on others (by analogy with empirical Bayes methods; see, e.g., Morris, 1983). We have done some work on graphical model checking using posterior predictive checks for problems with missing and latent data (Gelman et al., 2005). Perhaps these methods can be connected to the ideas of Thompson and Geyer (2005) for fuzzy p -values in latent variable problems.

p -VALUES AND u -VALUES

Finally, let me note that the classical p -value, defined from a point null hypothesis, can be generalized in various ways when moving to p -values that depend on unknown parameters in composite hypotheses. In the simplest setting, the classical p -value has two properties:

- (i) It is the probability that a test statistic in the reference distribution exceeds its value in the data [$p.\text{value}(y) = \Pr(T(y.\text{rep}) > T(y))$].
- (ii) It is a function of data, $p.\text{value}(y)$, that is uniformly distributed under the reference distribution [i.e., $p.\text{value}(y.\text{rep}) \sim U(0, 1)$].

With a composite null hypothesis characterized by an unknown parameter θ , the two properties of the p -values cannot both, in general, hold. From the Bayesian perspective, as noted above, the p -value can be defined by generalizing (i) to obtain: posterior $p.\text{value}(y) = \Pr(T(y.\text{rep}, \theta) > T(y, \theta) | y)$. Instead, one can generalize (ii) and define a p -value as