Statistical Science
2005, Vol. 20, No. 4, 372–374
DOI 10.1214/088342305000000377

Comment: Fuzzy and Randomized Confidence Intervals and P-Values

Roger L. Berger and George Casella

1. INTRODUCTION

We thank Professor Geyer and Professor Meeden for their thought-provoking article. We hope to also be thought-provoking in response, for we pretty much disagree with their position.

The fuzzy procedures proposed by the authors result from examining the test function, $\phi(x, \alpha, \theta)$ in three different ways, as a function of each of the three variables. This is an interesting exercise, which has not been done before in this way, and the authors are to be commended for their innovation. However, we think the resulting procedures will be of limited practical interest.

The authors start with the belief that discontinuous coverage probability functions are somehow inherently bad, saying that they "perform badly" and "behave very badly," and refer to their properties as "flaws." The new fuzzy procedures eliminate these flaws by having coverage probabilities that are exactly equal to $1 - \alpha$ and test sizes that are exactly equal to α. However, these flaws are merely the properties of discrete data, showing us the limit of the possible inference. To go beyond the inherent limitations of the data is to base inference on mathematical fictions. In particular, oscillations are just a feature of coverage probability with discrete data, and there is no principle that says coverage probability functions should be continuous. Although it is probably good if a coverage probability function stays close to $1-\alpha$, so the intervals of Blaker (2000) might be preferred to those of Clopper and Pearson, we do not see a need (or a way!) to eliminate discontinuities.

Procedures already exist that have coverage probabilities exactly equal to $1 - \alpha$ and sizes equal to α; they are classical randomized procedures. However, randomized procedures are unpalatable in actual data analysis, because, as the authors state, users object to a procedure that can give different answers for the exact same data. Unfortunately, the fuzzy procedures proposed by these authors are closely related to, in some cases almost indistinguishable from, randomized procedures. As such, we believe they will be equally unpalatable for practical inference. The fuzzy procedures do not give different answers for the same data; instead they give a single, different, harder to interpret answer for a given set of data. When the fuzzy procedures are used to produce confidence intervals and P-values in the usual sense, they simply result in classical randomized procedures.

2. WHAT IS FUZZY?

The description of the fuzzy confidence sets and Figure 2 are interesting. Instead of stating an interval of θ values as the inference, as classical nonrandomized and randomized confidence intervals do, the inference from a fuzzy confidence interval is a function, examples of which are shown in Figure 2. They are somewhat appealing, with their representation of the uncertainty of the inclusion probabilities of the endpoints, but will these functions be useful to or interpretable by researchers?

In classical confidence intervals there is one kind of uncertainty quantified by the confidence coefficient, $1 - \alpha$. This still is present in fuzzy intervals, but in fuzzy intervals there is a second uncertainty about the endpoints of the interval, represented by the ascending and descending portions of the functions in Figure 2. We all know the difficulty in teaching students the correct interpretation of the uncertainty quantified in $1 - \alpha$. How much more open to misinterpretation will be the uncertainty in the endpoints? The authors say that randomization is a "notoriously tricky concept," but are "partial coverage," "degree of membership" and "degree of compatibility" any less tricky?

To overcome this difficulty in interpretation, one can use the fuzzy interval to produce a realized randomized interval as described in Section 2.1. However, a realized randomized interval is just a classical randomized

Roger Berger is Professor and Chair, Department of Mathematical Sciences and Applied Computing, Arizona State University at the West campus, Phoenix, Arizona 85069, USA (e-mail: Roger.Berger@asu.edu). George Casella is Distinguished Professor and Chair, Department of Statistics, University of Florida, Gainesville, Florida 32611, USA (e-mail: casella@stat.ufl.edu).