*Statistical Science*
2005, Vol. 20, No. 4, 384–387
DOI 10.1214/088342305000000359


# Rejoinder: Fuzzy and Randomized Confidence Intervals and $P$-Values

**Charles J. Geyer and Glen D. Meeden**

We thank all the discussants for their insightful comments. We enjoyed reading the historical background they supplied, were pleased by their new ideas for fuzzy procedures and were provoked to produce better arguments for our ideas (which is what comments are supposed to do).

## 1. NEW FUZZY PROCEDURES

We think the most illuminating aspect of the comments is the new fuzzy (or abstract randomized) procedures they propose.

### 1.1 Two New Binomial Fuzzy Confidence Intervals

Agresti and Gottard propose an equal-tailed fuzzy interval they attribute to Stevens (1950), although, of course, the notion of a *fuzzy* confidence interval was not exactly what Stevens proposed. This is the fuzzy confidence interval with membership function given by (1.1b) of our article, where $\phi$ is the critical function of the equal-tailed randomized test.

Brown, Cai and DasGupta propose a fuzzy interval they attribute to Pratt (1961), although, of course, the notion of a *fuzzy* confidence interval was not exactly what Pratt proposed. This is the fuzzy confidence interval with membership function given by (1.1b), where $\phi(\cdot, \alpha, \theta)$ is the critical function of the most powerful randomized simple-versus-simple test with null hypothesis that the data are $\text{Binomial}(n, \theta)$ and alternative hypothesis that the data have the discrete uniform distribution on $\{0, \ldots, n\}$.

Figure 1 herein shows these two new fuzzy intervals along with the UMPU fuzzy intervals we proposed. Clearer and larger figures for more values of $x$ are given on the web (`www.stat.umn.edu/geyer/fuzz`). From the figure it can be seen that the Pratt (Brown–Cai–DasGupta) intervals are not unimodal, a point noted by Pratt (1961) and by Brown, Cai, and DasGupta in their comments. These fuzzy intervals arise from an optimality argument we think shows a fundamental misunderstanding of fuzzy confidence intervals (which, of course, we cannot anachronistically blame Pratt for). From our point of view, what they actually do is optimally test against an alternative (discrete uniform) that we cannot imagine will ever be of interest in applications. Nevertheless, we say the more the merrier. If one likes these fuzzy intervals, then use them.

The equal-tailed (Agresti–Gottard) tests are more reasonable. There is little practical difference between their proposal and ours. As they say, their intervals look more reasonable for $x$ in the middle of the range and ours look more reasonable elsewhere, but good frequentists cannot think this way (however natural it may be), since any frequentist property depends on averaging over all $x$.

### 1.2 UMPU Fuzzy Intervals Defended

Define the *coverage* at a point $\theta'$ of a fuzzy confidence interval (1.1b) when $\theta$ is the true parameter value to be

$$c(\theta, \theta') = E_\theta\{1 - \phi(X, \alpha, \theta')\}. \tag{1}$$

When $\theta = \theta'$, this is the left-hand side of (1.3) in our article.

The UMPU properties transferred to the language of confidence intervals are as follows:

(i) The interval is exact, that is,

$$c(\theta, \theta) = 1 - \alpha \quad \text{for all } \theta.$$

(ii) The interval has higher coverage for the true unknown $\theta$ than any other $\theta$, that is,

$$c(\theta, \theta) \geq c(\theta, \theta') \quad \text{for all } \theta \text{ and } \theta'.$$

(iii) Subject to the constraints (i) and (ii), the interval has the lowest possible coverage for all nontrue $\theta$, that is,

$$c(\theta, \theta') \leq \tilde{c}(\theta, \theta') \quad \text{whenever } \theta' \neq \theta,$$

where $\tilde{c}$ is the coverage for any other fuzzy confidence interval satisfying (i) and (ii) with $c$ replaced by $\tilde{c}$.

*Charles J. Geyer and Glen D. Meeden are Professors, School of Statistics, University of Minnesota, Minneapolis, Minnesota 55455, USA (e-mail: charlie@stat.umn.edu; glen@stat.umn.edu).*

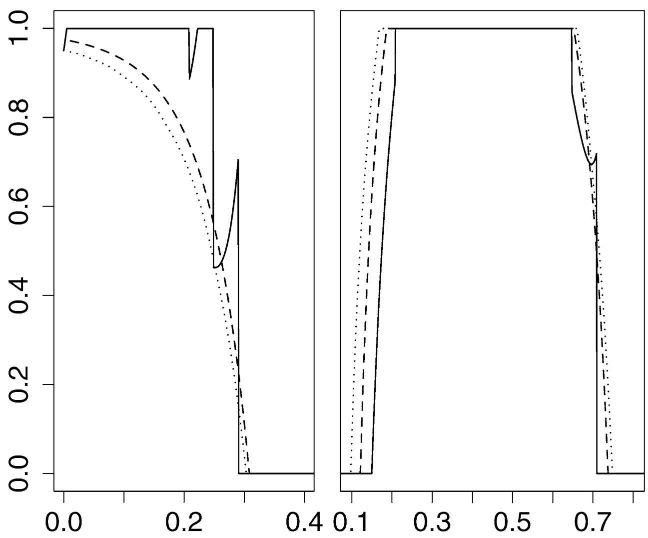


FIG. 1. *Fuzzy confidence intervals*: 95% *fuzzy confidence intervals for the binomial distribution*, $n = 10$. *Left panel*, $x = 0$; *right panel*, $x = 4$. *The solid curve is Pratt* (*Brown–Cai–DasGupta*), *the dashed curve is equal-tailed* (*Agresti–Gottard*) *and the dotted curve is UMPU* (*Geyer–Meeden*).

We should perhaps have written this out in detail in our article.

One may or may not be convinced that (ii) and (iii) are desirable properties, but temporarily accept them for the sake of argument. From this point of view the focus on *length* of the fuzzy interval (the side-to-side distance at various levels) seen in Brown, Cai and DasGupta and to a lesser degree in Agresti and Gottard is wrong. Property (iii) says that *height* (up–down distance) is the important criterion, and the UMPU intervals minimize it, subject to (i) and (ii), *uniformly* in $\theta$.

Since we have exact intervals which have a long-standing and well accepted optimality property, we see no need to introduce another optimality criterion, in particular, the one that arises from the quasi-Bayesian averaging over $\theta$ found in Pratt (1961) and in Brown, Cai and DasGupta (2001, and their comments here).

The criticism by Agresti and Gottard that our UMPU intervals go to $1 - \alpha$ as $\theta$ goes to zero when $x = 0$ is "unappealing behavior" is clearly wrong. All exact fuzzy confidence intervals, including their proposal, must have this property, as the left panel in Figure 1 (herein) shows. It is a simple consequence of the degeneracy of the binomial distribution as the parameter goes to the boundary. Moreover, their criticism misunderstands fuzzy confidence intervals. The fuzzy interval is saying "I don't *need* to go all the way up to 1.0 to get 95% coverage when $x = 0$."

The criticism by Agresti and Gottard that our UMPU intervals have support that is wider than Clopper–Pearson intervals again focuses on the wrong criterion (length instead of height) and misunderstands (we say) what fuzzy confidence intervals do.

A final point for the defense not mentioned by any discussant is the duality of tests and confidence intervals. If UMPU is a good idea for tests, then we hold it must also be a good idea for confidence intervals and vice versa. To hold otherwise is incoherent, but none of the discussants attacked UMPU tests.

### 1.3 New Fuzzy Procedures Unrelated to the Binomial Distribution

The most important aspect of fuzzy confidence intervals and $P$-values is that they are applicable much more generally than our article. Thompson's comments give an important application more distant from the binomial distribution than we dreamed. Aside from the scientific importance of this application, we think it is illuminating that the notion of fuzzy $P$-values leads to great simplification of a very difficult problem that had already been much studied (Thompson and Basu, 2003, and the other references cited in Thompson's comments). Readers who can get past the quibbling about the binomial distribution may see the potential of fuzzy confidence intervals and $P$-values, and find similarly original applications to new problems.

A much simpler application to new problems is to classical nonparametric tests in Geyer (2005).

### 1.4 Simultaneous Confidence Intervals and Multiple Tests

Brown, Cai and DasGupta give an interesting proposal for simultaneous fuzzy confidence intervals (their Section 4), which they call "abstract randomized confidence intervals," missing our point that there is no *unique* way to associate randomized confidence intervals with their Figure 3. Aside from this quibble about terminology, we like their idea.

Thompson does not mention it in her comments, but the technical report (Thompson and Geyer, 2005) does apply fuzzy $P$-values to multiple testing.

## 2. OTHER FUZZY STATISTICS

Some fuzzy set theorists, for example, Filzmoser and Viertl (2004), have imported fuzzy set theory into statistics, right at the beginning making *data* fuzzy. One "observes" random data, which are also fuzzy. Technically, the data are random functions (the membership functions of random fuzzy sets). This fuzziness then propagates through the whole analysis, resulting in fuzzy confidence intervals and fuzzy $P$-values, where the fuzziness has arisen from the initial fuzziness of the data and not through the ideas in our article.

In a different direction, Singpurwalla and Booker (2004) have proposed a model which allows the incorporation of fuzzy membership functions into subjective Bayesian inference. They do not give them a probabilistic interpretation.

Gelman suggests other Bayesian connections, asking about the relationship of Thompson's fuzzy $P$-values and posterior predictive $P$-values. What Thompson calls unobservable latent variables, a Bayesian would call parameters. Hence what Thompson calls the conditional distribution of these under the null hypothesis, a Bayesian would call the posterior (under the null hypothesis)? Whether a Bayesian would then go on to calculate the posterior distribution of the particular function of "parameters" that Thompson calls a fuzzy $P$-value and what interpretation the Bayesian would give it, we do not know. Has Thompson made an inadvertent contribution to Bayesian inference?

## 3. FISHER AND $P$-VALUES

To answer the question asked by Berger and Casella, we do follow Fisher (and Christensen, 2005) and want to directly interpret $P$-values, and no, this does not mean we think the FDA is "flipping coins," but neither do we accept that the FDA makes *statistical decisions* in the sense of statistical theory. Humans make the decisions. The $P$-values influence those decisions, but so do other scientific, political, ethical, social and business issues. This is what we meant by the "decisions" being merely metaphorical. We think most experienced users have some sense of what $P$-values mean that is independent of stories about metaphorical decisions. This sense is already somewhat vague, and the fact that a fuzzy $P$-value is smeared out over a (usually narrow) range makes it no less useful than conventional $P$-values. The lack of questions about this interpretation from the audiences for several talks on the subject confirms our views.

## 4. FUZZY IS TOO HARD?

One theme that runs through (to varying degrees) the comments of Agresti and Gottard, Berger and Casella and Brown, Cai and DasGupta is that our "fuzzy" procedures are not really new since some "abstract randomized" confidence intervals were described 50 years ago, by Stevens (1950), Blyth and Hutchinson (1960), Pratt (1961) and others mentioned in the comments. Moreover, since whatever promise Neyman may have thought they had was not achieved, it follows that our ideas must be nonstarters. We do not mind the accusations of nonoriginality, but we do object to fuzzy procedures not being given a fair hearing on their own terms.

A related theme in the comments is that introducing fuzzy procedures makes an already difficult area even more difficult. On the contrary, we believe that introducing some fuzzy terminology makes these procedures easier to understand, not harder. The difficulty seen by the commenters results from their refusal to take fuzzy seriously and their insistence on replacing it with abstract randomized.

The simplicity of conventional confidence intervals is an illusion fostered by the lack of understanding of users. If users were required to make a plot like Figure 1 of our article (or Figure 1 of Brown, Cai and DasGupta's comments or analogous figures cited in our article) instead of just claiming to have a "95% confidence interval," would they still think conventional procedures are simple? Yet such plots accurately describe what is being done. If not understood, then users simply do not understand what they are doing.

A fuzzy set is a simple concept that can be understood directly from a picture and will be obvious to any student who ever wanted *partial credit* on a test. All we are doing is introducing this notion of partial credit into confidence intervals. Like partial credit on a test, it does not involve any notion of randomness, abstract or otherwise, in its interpretation. You get full credit if the true parameter value $\theta$ is in the core of your fuzzy confidence interval and only partial credit if $\theta$ is elsewhere in its support. Of course, at the end we introduce some randomness: the coverage is the average amount of credit we get (averaged over all $x$).

We claim this view of fuzzy confidence intervals adds hardly any complication to the confidence interval notion. We claim even more: this view simplifies the explanation of conventional confidence intervals. Coverage is not like probability; it is like credit on a test question. Confidence is not like probability; it is like credit on a test averaged over all questions. The difficulty naive users have with confidence intervals is notorious. They confuse confidence with probability, but is this any surprise when they are given no theoretical notions other than probability? Having fuzzy as an alternative to probability can help make the distinction.

Of course all of this is alien to statisticians who are used to explaining everything in terms of probability, but what happens when you take a fuzzy confidence interval and replace it with an abstract randomized interval? You lose the simple "partial credit" interpretation and make the subject too complicated for ordinary users. We claim most of the critical comments are not about our proposal, but about a vaguely related, much more complicated and much less useful notion.

We developed fuzzy confidence intervals to be a useful replacement for randomized confidence intervals (realized or abstract) or conventional "crisp" confidence intervals (when the data are discrete). Fuzzy confidence intervals are a straightforward method that actually does what it claims to do, and one does not need a Ph.D. in statistics to understand it. This change in point of view and terminology should make confidence intervals easier for users to interpret correctly. Like Jourdain, in Moliere's *Le Bourgeois Gentilhomme*, who was pleased to learn he had been speaking prose all his life, we statisticians should be pleased to learn that we have been using fuzzy membership functions all our careers.

## REFERENCES

BLYTH, C. R. and HUTCHINSON, D. W. (1960). Table of Neyman—shortest unbiased confidence intervals for the binomial parameter. *Biometrika* **47** 381–391.

BROWN, L. D., CAI, T. T. and DASGUPTA, A. (2001). Interval estimation for a binomial proportion (with discussion). *Statist. Sci.* **16** 101–133.

CHRISTENSEN, R. (2005). Testing Fisher, Neyman, Pearson, and Bayes. *Amer. Statist.* **59** 121–126.

FILZMOSER, P. and VIERTL, R. (2004). Testing hypotheses with fuzzy data: The fuzzy $p$-value. *Metrika* **59** 21–29.

GEYER, C. J. (2005). Fuzzy $P$-values and ties in nonparametric tests. Available at www.stat.umn.edu/geyer/fuzz.

PRATT, J. W. (1961). Length of confidence intervals. *J. Amer. Statist. Assoc.* **56** 549–567.

SINGPURWALLA, N. D. and BOOKER, J. M. (2004). Membership functions and probability measures of fuzzy sets (with discussion). *J. Amer. Statist. Assoc.* **99** 867–889.

STEVENS, W. L. (1950). Fiducial limits of the parameter of a discontinuous distribution. *Biometrika* **37** 117–129.

THOMPSON, E. A. and BASU, S. (2003). Genome sharing in large pedigrees: Multiple imputation of *ibd* for linkage detection. *Human Heredity* **56** 119–125.

THOMPSON, E. A. and GEYER, C. J. (2005). Fuzzy $p$-values in latent variable problems. Technical Report 481, Dept. Statistics, Univ. Washington. Available at www.stat.washington.edu/www/research/reports/2005/tr481.pdf.