Fuzzy Confidence Intervals and P-values

Charles J. Geyer and Glen D. Meeden

November 14, 2002

Abstract. The optimal hypothesis tests for the binomial distribution and some other discrete distributions are uniformly most powerful (UMP) one-tailed and UMP unbiased (UMPU) two-tailed randomized tests. Conventional confidence intervals are not dual to randomized tests and perform badly on discrete data at small and moderate sample sizes. We introduce a new confidence interval notion, called fuzzy confidence intervals, that is dual to and inherits the optimality of UMP and UMPU tests. The analogous concept for P-values, called fuzzy P-values, also inherits the same optimality.

Key words and phrases: Confidence interval, P-value, hypothesis test, uniformly most powerful unbiased (UMP and UMPU), fuzzy set theory, randomized test.

1 Introduction

1.1 Conventional Confidence Intervals are Bad

The main point of this paper is that conventional confidence intervals, which we also call "crisp" confidence intervals for reasons that will presently become apparent, are a really bad idea for discrete data.

The badness of conventional confidence intervals has long been recognized, although seldom said so blatantly. A recent article (Brown, Cai, and DasGupta, 2001) reviews crisp confidence intervals for binomial models. The authors and discussants of this paper do recommend some crisp confidence intervals (not all recommending the same intervals), and the crisp confidence intervals they recommend are indeed better than the intervals they abhor (for some definitions of "better"). But even the best crisp confidence intervals behave very badly. The actual achieved confidence level oscillates wildly as a function of both the true unknown parameter value and the sample size. See our Figure 1, Figures 1–5, 10, and 11 in Brown, et al. (2001), Figures 4 and 5 in Agresti and Coul (1998), and Figure 1 in Casella (2001).

It is important that the reader not get lost in the plethora of crisp intervals that have been discussed in the literature. All crisp intervals for discrete data must exhibit wild oscillation similar to that shown in Figure 1. The fundamental

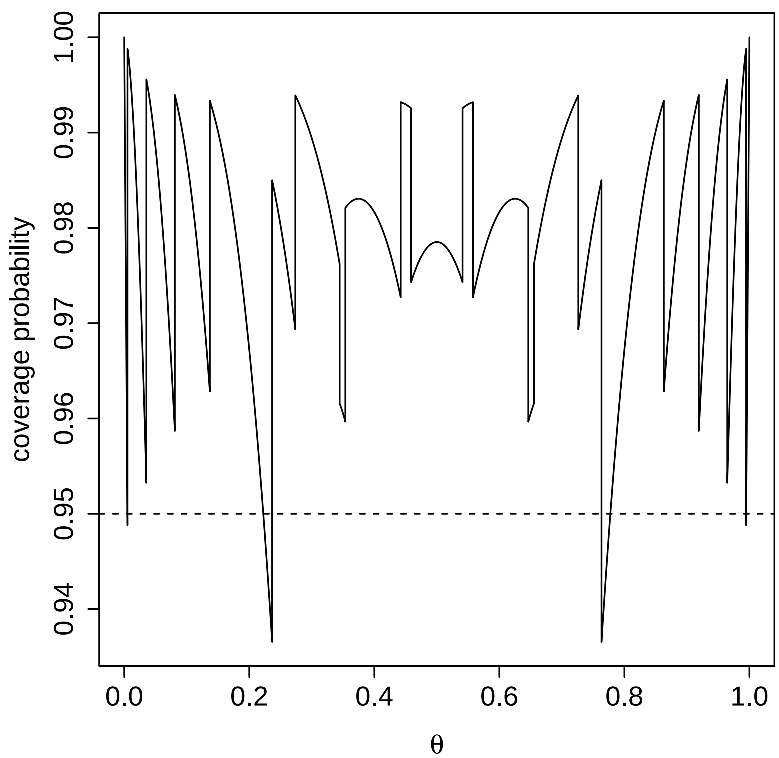

Figure 1: Coverage probability of the nominal 95% confidence interval for the binomial distribution with sample size $n = 10$ calculated by the function `prop.test` in the R statistical computing language (Ihaka and Gentleman, 1996). This is the Wilson (also called, score) interval with continuity correction and modifications when $x = 0$ or $x = n$. The dashed line is the nominal level. The solid line is the coverage probability of the interval as a function of the success probability θ.

reason is discreteness. When the data x are discrete, so are the endpoints $l(x)$ and $u(x)$ of possible crisp confidence intervals. As the parameter θ passes from just inside to just outside a possible confidence interval $\bigl(l(x), u(x)\bigr)$, the coverage probability drops discontinuously by $\mathrm{pr}_\theta(X = x)$. This flaw is unavoidable, an irreconcilable conflict between crisp confidence intervals and discrete data.

As everyone knows, standard asymptotic theory says that as the sample size goes to infinity the oscillations get smaller in small neighborhoods of one fixed parameter value θ in the interior of the parameter space. For the binomial distribution, this means the oscillations get smaller for θ not near zero or one. But the oscillations remain large for shockingly large sample sizes (Brown, et al., 2001) and the oscillations remain large for all sample sizes for θ near zero and one.

The behavior of conventional confidence intervals for discrete data is so bad that it leads one to question the whole idea. From our point of view, all of the many methods for producing crisp confidence intervals for the binomial distribution are attempts to make a silk purse out of a sow's ear. The inherent flaws of the crisp confidence interval idea make any solution bad. Arguments about slightly less terrible or slightly more terrible performance of various crisp intervals ignore the fundamental badness of all of them.

We started to question the whole idea of crisp confidence intervals when we considered the question of what is a "randomized confidence interval." Everyone knows that the testing problem for discrete models was solved long ago by the introduction of randomized tests. For the binomial distribution and many other discrete distributions there exist uniformly most powerful (UMP) one-tailed tests and UMP unbiased (UMPU) two-tailed tests (Lehmann, 1959, Chapters 3 and 4). These tests are optimal procedures. Tests and confidence intervals are dual notions. Why doesn't the duality save us here? What confidence interval notion is dual to randomized tests? Not, unfortunately, conventional confidence intervals.

Further consideration of the issue led us away from the desire for a "randomized" confidence interval. Randomized tests make beautiful theory, but in practice many people object to a procedure that can give different answers for the exact same data due to the randomization. We can keep the beautiful theory while ditching the arbitrariness of the randomization, by a simple change of viewpoint to what we call "fuzzy" concepts.

1.2 Fuzzy Decisions, Confidence Intervals, and P-values

Let ϕ be the critical function of a (randomized) test. It is a function from the sample space to the interval $[0, 1]$. Since it is a function on the sample space, it is usually written $\phi(X)$, but since the function also depends on the size of the test and the hypothesized value of the parameter, we prefer to write it $\phi(X, \alpha, \theta)$, where X is the data, α is the significance level (size), and θ is the hypothesized value of the parameter under the null hypothesis. A *randomized test* of size α rejects $H_0 : \theta = \theta_0$ when data x are observed with probability $\phi(x, \alpha, \theta_0)$. [The exact form of the critical function does not matter for the discussion in this section. The curious who have forgotten UMP and UMPU theory can look at equations (3.1) and (3.4) below].

Randomized tests have the annoying property that the reported decision ("accept H_0" or "reject H_0") is not a function of the observed data x alone. Two statisticians analyzing exactly the same data can report different results. For this reason we make a change that is theoretically trivial but of some practical importance. We think a statistician using a randomized test should just report the value $\phi(x, \alpha, \theta_0)$. We call this a *fuzzy test* and the reported value a *fuzzy decision*. A statistician preferring a classical randomized test can always generate his or her own Uniform$(0, 1)$ random variate U and utter "reject H_0" if $U < \phi(x, \alpha, \theta_0)$ and "accept H_0" otherwise.

Now we are ready to define two new concepts. For comparison we repeat the definition we just gave of fuzzy decision.

- For fixed α and θ_0, the function $x \mapsto \phi(x, \alpha, \theta_0)$ is the fuzzy decision function for the size α test of $H_0 : \theta = \theta_0$.
- For fixed x and α, the function $\theta \mapsto 1 - \phi(x, \alpha, \theta)$ is (the membership function of) the fuzzy confidence interval with coverage $1 - \alpha$.
- For fixed x and θ_0, the function $\alpha \mapsto \phi(x, \alpha, \theta_0)$ is (the cumulative distribution function of) the fuzzy P-value for the test of $H_0 : \theta = \theta_0$.

The fuzzy confidence interval $\theta \mapsto 1 - \phi(x, \alpha, \theta)$ is a function taking values between zero and one and is to be interpreted as a fuzzy set. We will say more about fuzzy sets below (Section 2.1). For now we will just say that the fuzzy confidence interval gives for each θ a number between zero and one that says the degree to which we should consider θ to be in the interval. Figure 2 shows a few fuzzy confidence intervals. The dashed line in the figure is the (graph of the membership function of) the fuzzy confidence interval for $n = 10$ and $x = 4$. It is zero for $\theta < 0.09775$ or $\theta > 0.74863$, meaning such points are definitely not in the fuzzy interval. It is one for $0.16875 < \theta < 0.66045$, meaning such points are definitely in the confidence interval. Parameter values in the "edges" of the fuzzy interval ($0.09775 < \theta < 0.16875$ and $0.66045 < \theta < 0.74863$) are ambiguous, neither definitely in nor definitely out. The rise and fall at the edges is fairly steep. Most points are definitely in or definitely out, so fuzzy intervals are not too different from crisp intervals. The amount of fuzziness is fairly small (and is smaller still for larger sample sizes).

The test having critical function ϕ is *exact* if

$$E_\theta\{\phi(X, \alpha, \theta)\} = \alpha, \qquad \theta \in \Theta,\ 0 \le \alpha \le 1. \tag{1.1}$$

Note that this trivially implies

$$E_\theta\{1 - \phi(X, \alpha, \theta)\} = 1 - \alpha, \qquad \theta \in \Theta,\ \ 0 \le \alpha \le 1. \tag{1.2}$$

The left hand side of (1.2) is the coverage probability of the fuzzy confidence interval. Hence the fuzzy confidence interval inherits the exactness of the corresponding test. Since UMP and UMPU tests are exact, so are the corresponding fuzzy confidence intervals.

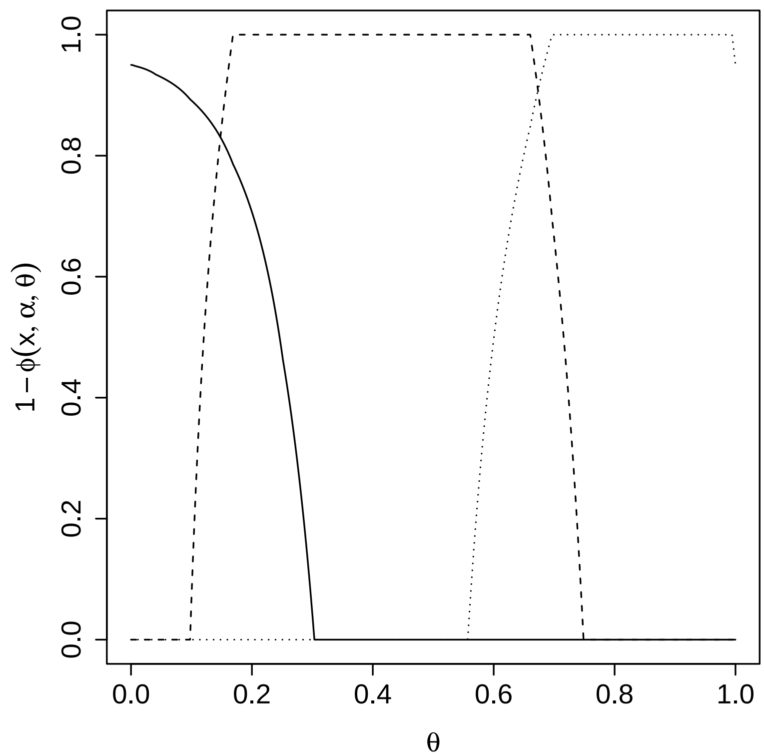

Figure 2: Fuzzy confidence intervals for binomial data with sample size $n = 10$, confidence level $1 - \alpha = 0.95$ and observed data $x = 0$ (solid curve) $x = 4$ (dashed curve) and $x = 9$ (dotted curve). Note that the $x = 0$ curve starts at $1 - \alpha$ at $\theta = 0$ and the $x = 9$ curve ends at $1 - \alpha$ at $\theta = 1$ (Section 3.2 below explains this behavior). The parameter θ is the probability of success.

The fuzzy P-value $\alpha \mapsto \phi(x, \alpha, \theta_0)$ is a nondecreasing continuous function that goes from zero to one as α goes from zero to one. We interpret it as the distribution function of a continuous random variable P. By definition of distribution function,

$$\mathrm{pr}_\theta\{P \le \alpha \mid X\} = \phi(X, \alpha, \theta),$$

for all $\theta \in \Theta$ and $0 \le \alpha \le 1$. Hence by iterated conditional expectation

$$\begin{aligned} \mathrm{pr}_\theta\{P \le \alpha\} &= E_\theta\{\mathrm{pr}_\theta\{P \le \alpha \mid X\}\} \\ &= E_\theta\{\phi(X, \alpha, \theta)\} \\ &= \alpha, \end{aligned} \tag{1.3}$$

for all $\theta \in \Theta$ and $0 \le \alpha \le 1$. The left hand side of (1.3) is the significance level of the hypothesis test using the fuzzy P-value. Hence the fuzzy P-value inherits the exactness of the corresponding test. Since UMP and UMPU tests are exact, so are the corresponding fuzzy P-values.

The derivative

$$\alpha \mapsto \frac{\partial}{\partial \alpha} \phi(x, \alpha, \theta_0)$$

is the probability density function of the continuous random variable that is the fuzzy P-value. Figure 3 shows the probability density function of a fuzzy P-value. As we show below (Section 3.1), every probability density function of a fuzzy P-value corresponding to a UMP or UMPU test is a step function.

All fuzzy P-values for UMP tests and some for UMPU tests have densities with just one step, that is, they are uniformly distributed. In general, the fuzzy P-values in the light tail will be uniform and those in the heavy tail will be non-uniform. For example, for the same null hypothesis $\theta_0 = 0.7$, the same two-tailed alternative, and same sample size $n = 10$ as in Figure 3, if we observe $x = 3$, then the fuzzy P-value is $\text{Uniform}(0.0043, 0.0253)$, and if we observe $x = 2$, then the fuzzy P-value is $\text{Uniform}(0.0004, 0.0043)$.

2 Fuzziness and Randomness

2.1 Fuzzy Set Theory

So why are we calling these concepts *fuzzy*? Obviously, we have taken the term from fuzzy set theory, but what does that field have to do with randomized tests? Actually, we use none of the theory from fuzzy set theory, just concepts and terminology, which can be found in the most elementary of introductions to the subject (Klir, St. Clair, and Yuan, 1997).

A *fuzzy set* A in a space S is characterized by its *membership function*, which is a map $I_A : S \to [0, 1]$. The value $I_A(x)$ is the "degree of membership" of the point x in the fuzzy set A or the "degree of compatibility … with concept represented by the fuzzy set" (Klir, et al., 1997, p. 75). The idea is that we are uncertain about whether x is in or out of the set A. The value $I_A(x)$ represents

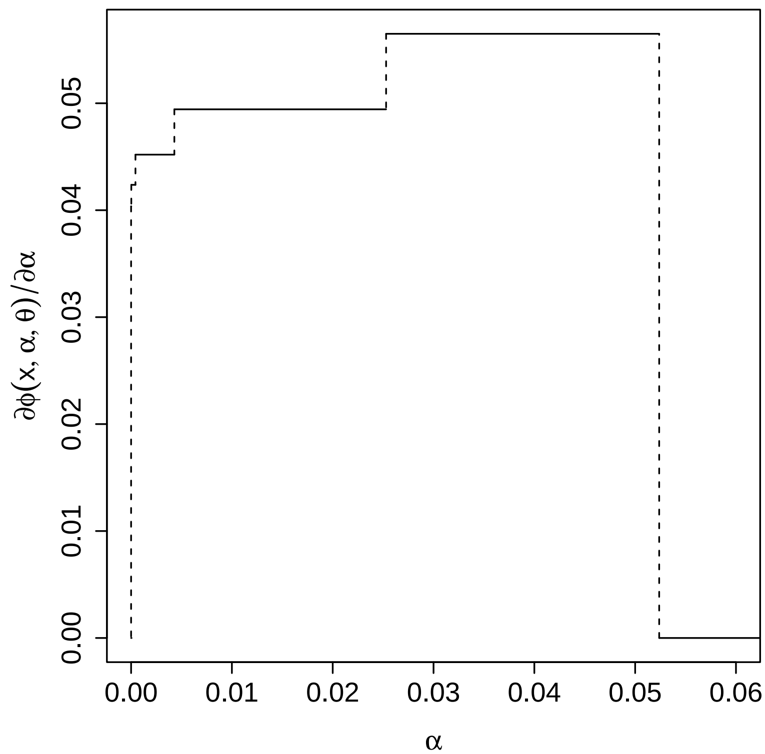

Figure 3: The Density of the Fuzzy P-value for the UMPU (two-tailed) test with binomial data $x = 10$ and $n = 10$ and null hypothesis $\theta = 0.7$. The plotted step function has five steps. One is too small too see, at height 0.0404 on the interval $2 \times 10^{-5} < \alpha < 4.3 \times 10^{-4}$.

how much we think x is in the fuzzy set A. The closer $I_A(x)$ is to 1.0, the more we think x is in A. The closer $I_A(x)$ is to 0.0, the more we think x is not in A.

Statisticians, especially subjective Bayesians are at this point likely to jump to the conclusion that fuzzy set theory can be replaced by probability theory. Is it not subjective Bayes dogma that probability is the only coherent measure of uncertainty? The fuzzy set crowd must either be incoherent (in the technical sense meant by Bayes theorists) or they must be just redoing probability in some form.

Readers may rest assured that fuzzy set theorists are not redoing probability in disguised form. From the origins of their subject more than 30 years ago they have taken great pains to distinguish their subject from probability theory, and the least acquaintance with the manipulations done in fuzzy set theory shows no resemblance at all to those of probability theory. We stress this point because it is so natural for statisticians (certain statisticians anyway) to try to find priors or posteriors somewhere in our discussion. Let us assure all readers that this paper is entirely non-Bayesian and that whatever fuzzy confidence intervals may be, they aren't Bayesian. We don't say this because we're anti-Bayes. We've looked for the Bayes angle and satisfied ourselves that it just isn't there. This shouldn't be surprising. There are few less Bayesian areas of statistics than confidence intervals and P-values. Making them fuzzy doesn't make them Bayesian.

A fuzzy set whose membership function actually only takes the values zero or one is called *crisp*. For a crisp set, the membership function I_A is the same thing as the indicator function of an ordinary set A. Thus "crisp" is just the fuzzy set theory way of saying "ordinary," and "membership function" is the fuzzy set theory way of saying "indicator function." We have adopted this terminology, calling conventional confidence intervals *crisp confidence intervals*.

Note that this makes conventional confidence intervals a special case of fuzzy confidence intervals (the fuzzy intervals that just happen to be crisp). Thus our fuzzy theory is a generalization of current theory. It includes all current results. In particular, fuzzy confidence intervals based on UMP and UMPU tests for *continuous* data are automatically crisp because those UMP and UMPU tests are not randomized. So our theory only says new things about discrete data. For continuous data, it's the same old story.

A decision for a test of statistical hypotheses is an element of the two-element set

$$D = \{\text{reject } H_0, \text{accept } H_0\}.$$

A fuzzy decision is just the fuzzy subset of D with membership function that has the value $\phi(x, \alpha, \theta)$ at the point "reject H_0" and the value $1 - \phi(x, \alpha, \theta)$ at the point "accept H_0." So the notion of fuzzy decisions fits well into fuzzy set theory.

We concede that "fuzzy P-value" is perhaps a misnomer for the the concept we have defined and given that name. We hope nobody objects because it does lend terminological similarity to three very closely related concepts (fuzzy decisions, confidence intervals, and P-values). In fact, there is nothing less fuzzy than a random variable. Probability theory and fuzzy set theory have an empty intersection.

But other nomenclatural ideas don't seem appropriate either. To call our new P-values "randomized" or "random" suggests that they are, like the decision of a randomized test, *outcomes* or *realized values* of a random variable rather than *random variables* themselves. Other possible terms, such as "random-variable P-values" or "P-values that are random variables" are not only unwieldy but also confusing.

Thus we stick with "fuzzy P-value," misnomer though it may be.

2.2 Randomized Analogs of Fuzzy Concepts

To each fuzzy concept in the trio of decisions, confidence intervals, and P-values, there is an analogous randomized concept. For decisions the correspondence is trivial and obvious. The other two need a bit of explanation.

2.2.1 P-Values

The important theoretical distinction made at the end of the preceding section between a fuzzy P-value as a random variable in the theoretical sense (a function on a sample space) as opposed to an observed value of a random variable (an ordinary number) tells us what a *randomized P-value* is: a random realization from the distribution of the fuzzy P-value. Let P denote a simulated random variate having the distribution of the fuzzy P-value. Then the test that rejects H_0 when $P \leq \alpha$ is the traditional randomized test.

We show below (Section 3.1) that in the UMP case the fuzzy P-value is uniformly distributed on an interval (3.3) and in the UMPU case the fuzzy P-value is a continuous random variable having piecewise constant density (hence a mixture of uniforms). In either case, simulating a realization of the fuzzy P-value is trivial given a uniform random number generator.

2.2.2 Confidence Intervals

We need another bit of fuzzy set terminology. If I_A is the membership function of a fuzzy set A, the γ-*cut* of A (Klir, et al., 1997, Section 5.1) is the crisp set

$$^{\gamma}I_A = \{\, x : I_A(x) \geq \gamma \,\}.$$

Clearly, knowing all the γ-cuts for $0 \leq \gamma \leq 1$ tells us everything there is to know about the fuzzy set A.

Now let

$$I_x(\theta) = 1 - \phi(x, \alpha, \theta)$$

be the fuzzy confidence interval with coverage $1 - \alpha$ for observed data x. Then the corresponding *randomized confidence interval* is the W-cut $^{W}I_X$ of the fuzzy interval, where W is a Uniform$(0, 1)$ random variate.

By construction

$$\mathrm{pr}_\theta\{\theta \in {}^{W}I_X \mid X\} = E_\theta\{I_X(\theta) \mid X\} = 1 - \phi(X, \alpha, \theta).$$

So

$$\mathrm{pr}_\theta\{\theta \in {}^W I_X\} = E_\theta\{I_X(\theta)\} = 1 - \alpha$$

and the randomized confidence interval inherits exactness from the fuzzy confidence interval.

3 A Little Bit of Theory

3.1 Fuzzy P-Values for UMP and UMPU

3.1.1 UMP

Lehmann (1959, pp. 68–69) says for a one-parameter model with likelihood ratio monotone in the statistic $T(X)$ there exists a UMP test having null hypothesis $H_0 = \{\,\vartheta : \vartheta \le \theta\,\}$, alternative hypothesis $H_1 = \{\,\vartheta : \vartheta > \theta\,\}$, significance level α, and critical function ϕ defined by

$$\phi(x, \alpha, \theta) = \begin{cases} 1, & T(x) > C \\ \gamma, & T(x) = C \\ 0, & T(x) < C \end{cases} \tag{3.1}$$

where the constants γ and C are determined by

$$E_\theta\{\phi(X, \alpha, \theta)\} = \alpha.$$

As Lehmann says, the description of the analogous lower-tailed test is the same except that all inequalities are reversed.

The constant C is clearly any $(1-\alpha)$-th quantile of the distribution of $T(X)$ for the parameter value θ. If C is not an atom of this distribution, then the test is effectively not randomized and the value of γ is irrelevant. Otherwise

$$\gamma = \frac{\alpha - \mathrm{pr}_\theta\{T(X) > C\}}{\mathrm{pr}_\theta\{T(X) = C\}}. \tag{3.2}$$

In considering the distribution of the fuzzy P-value, we look at $\phi(x, \alpha, \theta)$ as a function of α for fixed x and θ, hence at (3.2) in the same way. Now $T(x)$ will be a $(1 - \alpha)$-th quantile if

$$\mathrm{pr}_\theta\{T(X) > T(x)\} \le \alpha \le \mathrm{pr}_\theta\{T(X) \ge T(x)\}.$$

Hence

$$\phi(x, \alpha, \theta) = \begin{cases} 0, & \mathrm{pr}_\theta\{T(X) \ge T(x)\} \ge \alpha \\ 1, & \mathrm{pr}_\theta\{T(X) > T(x)\} \le \alpha \\ \frac{\alpha - \mathrm{pr}_\theta\{T(X) > T(x)\}}{\mathrm{pr}_\theta\{T(X) = T(x)\}}, & \text{otherwise} \end{cases}$$

Since this is linear where not zero or one, it is the distribution function of a uniform random variable

$$P \sim \mathrm{Uniform}\big(\mathrm{pr}_\theta\{T(X) > T(x)\}, \mathrm{pr}_\theta\{T(X) \ge T(x)\}\big), \tag{3.3}$$

that is, the fuzzy P-value is uniformly distributed on the interval of values that you figure it should spread it over. Intuitively obvious, but we had to check to be sure. Please don't feel put upon by our belaboring the obvious, because the UMPU two-tailed case is not quite so obvious, and understanding the UMP one-tailed case helps.

3.1.2 UMPU

Lehmann (1959, pp. 126–127) says for a one-parameter exponential family model with canonical statistic $T(X)$ and canonical parameter θ there exists a UMPU test having null hypothesis $H_0 = \{\, \vartheta : \vartheta = \theta \,\}$, alternative hypothesis $H_1 = \{\, \vartheta : \vartheta \neq \theta \,\}$, significance level α, and critical function ϕ defined by

$$\phi(x, \alpha, \theta) = \begin{cases} 1, & T(x) < C_1 \\ \gamma_1, & T(x) = C_1 \\ 0, & C_1 < T(x) < C_2 \\ \gamma_2, & T(x) = C_2 \\ 1, & C_2 < T(x) \end{cases} \tag{3.4}$$

where $C_1 \leq C_2$ and the constants γ_1, γ_2, C_1, and C_2 are determined by

$$E_\theta\{\phi(X, \alpha, \theta)\} = \alpha \tag{3.5a}$$

$$E_\theta\{T(X)\phi(X, \alpha, \theta)\} = \alpha E_\theta\{T(X)\} \tag{3.5b}$$

If $C_1 = C_2 = C$ in (3.4) then $\gamma_1 = \gamma_2 = \gamma$ also. This occurs only in a very special case. Define

$$p = \mathrm{pr}_\theta\{T(X) = C\} \tag{3.6a}$$

$$\mu = E_\theta\{T(X)\} \tag{3.6b}$$

Then in order to satisfy (3.5a) and (3.5b) we must have

$$1 - (1 - \gamma)p = \alpha$$

$$\mu - C(1 - \gamma)p = \alpha\mu$$

which solved for γ and C gives

$$\gamma = 1 - \frac{1 - \alpha}{p} \tag{3.7a}$$

$$C = \mu \tag{3.7b}$$

Thus this special case occurs only when μ an atom of the distribution of $T(X)$ for the parameter value θ, and then only for very large significance levels: $\alpha > 1-p$. Hence this special case is of no practical importance (Lehmann takes no explicit notice of it), although it is of some computational importance to get every case right, no weird bogus results or crashes in unusual special cases.

Now turn to the general case, assume for a second that we have particular C_1 and C_2 that work for some x, α, and θ (we will see how to determine them presently). With μ still defined by (3.6b) and with the definitions

$$p_i = \mathrm{pr}_\theta\{T(X) = C_i\}, \qquad i = 1, 2 \tag{3.8a}$$

$$p_{12} = \mathrm{pr}_\theta\{C_1 < T(X) < C_2\} \tag{3.8b}$$

$$m_{12} = E_\theta\{T(X) I_{(C_1,C_2)}[T(X)]\} \tag{3.8c}$$

(3.5a) and (3.5b) become

$$1 - (1 - \gamma_1)p_1 - (1 - \gamma_2)p_2 - p_{12} = \alpha \tag{3.9a}$$

$$\mu - C_1(1 - \gamma_1)p_1 - C_2(1 - \gamma_2)p_2 - m_{12} = \alpha\mu \tag{3.9b}$$

which solved for γ_1 and γ_2 give

$$\gamma_1 = 1 - \frac{(1 - \alpha)(C_2 - \mu) + m_{12} - C_2 p_{12}}{p_1(C_2 - C_1)} \tag{3.10a}$$

$$\gamma_2 = 1 - \frac{(1 - \alpha)(\mu - C_1) - m_{12} + C_1 p_{12}}{p_2(C_2 - C_1)} \tag{3.10b}$$

Note that (3.10a) and (3.10b) are linear in α. They are valid over the range of α (if any) such that both equations give values between zero and one.

Now we turn to the determination of C_1 and C_2. We present an algorithm that determines $\phi(x, \alpha, \theta)$ for any discrete one-parameter exponential family with canonical statistic $T(X)$ for all values of x and α for one fixed value of θ.

1. Start with $\alpha = 1$.

   (a) If μ given by (3.6b) is an atom, then $\phi(x, \alpha, \theta)$ is given by (3.4) with $C_1 = C_2 = \mu$ and $\gamma_1 = \gamma_2 = \gamma$ given by (3.6a), (3.7a), and (3.7b) over the range of α such that (3.7a) is between zero and one.

   (b) If μ given by (3.6b) is not an atom, then choose C_1 and C_2 to be adjacent atoms such that $C_1 < \mu < C_2$ and $\phi(x, \alpha, \theta)$ is given by (3.4) with γ_1 and γ_2 given by (3.8a), (3.8b), (3.8c), (3.10a), and (3.10b) over the range of α such that both (3.10a) and (3.10b) are between zero and one. [Because C_1 and C_2 are adjacent atoms, $p_{12} = m_{12} = 0$, and $\alpha = 1$ gives $\gamma_1 = \gamma_2 = 1$, a valid solution.]

2. Start with the smallest α for which $\phi(x, \alpha, \theta)$ was determined in step 1 or a previous execution of step 2. At this point, either γ_1 or γ_2 or both is zero.

   (a) If γ_1 is zero, then decrease C_1 to the adjacent lower atom and set $\gamma_1 = 1$ [which does not change the value of $\phi(x, \alpha, \theta)$ for any x].

   (b) If γ_2 is zero, then increase C_2 to the adjacent higher atom and set $\gamma_2 = 1$ [which does not change the value of $\phi(x, \alpha, \theta)$ for any x].

(c) Now $\phi(x, \alpha, \theta)$ is given by (3.7a) with γ_1 and γ_2 given by (3.8a), (3.8b), (3.8c), (3.10a), and (3.10b) over the range of α such that both (3.10a) and (3.10b) are between zero and one [because of steps (a) and (b), both γ_i are now greater than zero, so α can be decreased].

3. Repeat step 2 until the whole range $0 \le \alpha \le 1$ is covered.

This algorithm is certainly unwieldy, but it does make clear that $\alpha \mapsto \phi(x, \alpha, \theta)$ is (1) continuous, (2) piecewise linear, (3) nondecreasing, and (4) onto $[0, 1]$. Hence it is the distribution function of a continuous random variable (our fuzzy P-value). Clearly, it is differentiable on each linear piece and the derivative is piecewise constant (a step function).

3.2 Endpoint Behavior

The UMPU test is not well defined when the null hypothesis is on the boundary of the parameter space. But equations (3.4), (3.5a), and (3.5b) still make sense and define a test. Since the probability and the expectation in those equations are continuous in θ this also characterizes the behavior as θ converges to a boundary point (which we need to know to calculate fuzzy confidence intervals, which involve all θ in the parameter space).

The following setup is general enough to tell us what happens in practical applications. Suppose X is the nonnegative-integer-valued canonical statistic of a one-parameter exponential family.

The densities in the family have the form

$$f_\theta(x) = \frac{1}{c(\theta)} e^{x\theta} \lambda(x), \tag{3.11}$$

where λ is some nonnegative function and

$$c(\theta) = \sum_{x=0}^{\infty} e^{x\theta} \lambda(x). \tag{3.12}$$

We assume $\lambda(0) > 0$ and $\lambda(1) > 0$.

The natural parameter space of the family is the set Θ of θ such that (3.12) is finite. Clearly, if $c(\psi) < \infty$, then $c(\theta) < \infty$ for all $\theta < \psi$. Thus Θ is either the whole real line or a semi-infinite interval extending to $-\infty$.

For every θ in the interior of Θ, the distribution with density f_θ has a moment generating function M_θ defined by

$$M_\theta(t) = E_\theta\{e^{tX}\} = \frac{c(\theta + t)}{c(\theta)},$$

and hence this distribution has moments of all orders, the mean and variance being given by derivatives at zero of the cumulant generating function $\log M_\theta$

$$\mu(\theta) = E_\theta(X) = \frac{d}{d\theta} \log c(\theta)$$

$$\sigma^2(\theta) = \operatorname{var}_\theta(X) = \frac{d^2}{d\theta^2} \log c(\theta)$$

Because of our assumption that $\lambda(0)$ and $\lambda(1)$ are strictly positive, $\sigma^2(\theta) = d\mu(\theta)/d\theta$ can never be zero. Hence μ is a strictly increasing continuous function that maps the interior of Θ to some open interval of the real line.

Since

$$\frac{f_\theta(x)}{f_\theta(0)} = e^{x\theta} \frac{\lambda(x)}{\lambda(0)}$$

the distribution clearly converges to the distribution concentrated at zero as $\theta \to -\infty$.

Now

$$\frac{\mu(\theta)}{f_\theta(1)} = \sum_{x=1}^{\infty} x e^{(x-1)\theta} \frac{\lambda(x)}{\lambda(1)}$$

goes to one by monotone convergence as $\theta \to -\infty$. Hence $\mu(\theta) \to 0$ as $\theta \to -\infty$. Thus μ is a diffeomorphism from the interior of Θ to some open interval of the real line, the lower endpoint of which is zero.

Now

$$\frac{\text{pr}_\theta\{X > 1\}}{f_\theta(1)} = \sum_{x=2}^{\infty} e^{(x-1)\theta} \frac{\lambda(x)}{\lambda(1)}$$

goes to zero by monotone convergence as $\theta \to -\infty$.

And these facts together imply

$$\begin{aligned} \text{pr}_\mu\{X = 0\} &= 1 - \mu + o(\mu) \\ \text{pr}_\mu\{X = 1\} &= \mu + o(\mu) \\ \text{pr}_\mu\{X > 1\} &= o(\mu) \end{aligned} \tag{3.13}$$

where $\mu = \mu(\theta)$ is the mean value parameter.

Now we claim that for small enough values of θ or μ we have $C_1 = 0$ and $C_2 = 1$ and the UMPU test is given by (3.10a) and (3.10b) with p_1 and p_2 given by (3.8a) and $p_{12} = m_{12} = 0$. Let's check. These equations give now

$$\begin{aligned} \gamma_1 &= 1 - \frac{(1-\alpha)(1-\mu)}{1 - \mu + o(\mu)} \\ \gamma_2 &= 1 - \frac{(1-\alpha)\mu}{\mu + o(\mu)} \end{aligned}$$

and clearly both converge to α as $\mu \to 0$ hence both are between zero and one for small enough θ or μ and hence define the UMPU test.

What happens at an upper bound (such as at n for the binomial) is clearly similar, since exponential families have behavior unchanged by linear transformations. (Actually our discussion above characterizes the behavior of any discrete exponential family. The only changes required if X is not integer-valued are purely notational.)

This explains the behavior of the fuzzy confidence intervals for the two x values nearest the boundaries in Figure 2. As $\theta \to 0$, the fuzzy confidence interval $1 - \phi(x, \alpha, \theta)$ converges to $1 - \alpha$ for $x = 0$ or $x = 1$ and converges to zero for all other x. And as $\theta \to 1$, the fuzzy confidence interval converges to $1 - \alpha$ for $x = n - 1$ or $x = n$ and converges to zero for all other x.

4 Discussion

This discussion is a work of fiction. The characters and events portrayed are fictional and any resemblance to real people or events is purely coincidental.

Two statisticians, Fred and Sally, enter stage right.

Fred. Do you think anyone will ever use this fuzzy nonsense?

Sally. Why not? It's the right thing with a capital R and a capital T, uniformly most wonderful, and so forth.

Fred. But it's crazy. An interval is an *interval* between *two* numbers. Not a picture!

Sally. That's just your brainwashing. You don't want to admit what you've been teaching students for decades is wrong. But if this fuzzy stuff had been invented along with randomized tests back in the 1930's and been included in the brainwashing you got in graduate school, then you would be perfectly happy with it.

Fred. But it's too complicated. People want simple results they can understand. They won't ever understand this fuzzy stuff.

Sally. Most people don't understand most things in statistics. That's why we teach statistics courses. The picture for a fuzzy confidence interval or fuzzy P-value is no more complicated than a histogram. You could teach it in elementary courses if you wanted to.

Most of statistics is more complicated than fuzzy confidence intervals and P-values. That doesn't stop us from teaching it.

Fred. What I mean is that it shouldn't need a picture to report a confidence interval or hypothesis test. That's too complicated.

Sally. This from somebody who won't do a regression without making a half dozen plots? People always drag out the "too complicated" or "too time consuming" argument at exactly the wrong time. It's silly to say a statistical procedure must be simple enough to do or explain in 60 seconds. The data may have taken years to collect. If the data are important, they are worth a proper analysis. Never mind how long it takes to do or explain.

Fred. Anyway, this stuff only applies to the binomial distribution.

Sally. No, it applies to any exponential family. There's the Poisson and negative binomial too. Even for multiparameter exponential families, a test about one parameter of interest is UMP or UMPU. Just use the one-parameter exponential family obtained by conditioning on the canonical statistics for the nuisance parameters. With that you get comparison of two independent binomials or two independent Poissons or two independent negative binomials. In large contingency tables, there isn't usually a single parameter of interest, but in two-by-two tables, there are the UMP and UMPU competitors of Fisher's exact test and McNemar's test. There exist, though I don't know whether they've been described, UMP and UMPU tests for a regression parameter in logistic or Poisson regression.

Anyway, there's nothing that says you can't use fuzzy confidence intervals and P-values whenever you have discrete data. We don't know how to extend the UMP and UMPU constructions outside of exponential families. But the idea of randomized tests and their associated fuzzy decisions, confidence intervals, and P-values is perfectly general. In principle, they can be applied to any discrete data. It's an interesting open research question, not a fundamental limitation.

Exeunt stage left. End of Act I.

References

Agresti, A. and Coul, B. J. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *Amer. Statist.* **52** 119–126.

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

Casella, G. (2001). Comment on Brown, Cai, and DasGupta (2001). *Statist. Sci.* **16** 120–122.

Ihaka, R. and Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. *J. Comput. Graph. Statist.* **5** 299–314.

Lehmann, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York. (2nd ed., Wiley, 1986; Springer, 1997).

Klir, G. J., St. Clair, U. H. and Yuan, B. (1997). *Fuzzy Set Theory: Foundations and Applications.* Prentice Hall PTR, Upper Saddle River, NJ.