Algorithm for UMP and UMPU Tests in Some Discrete Exponential Families

Charles J. Geyer

March 31, 2024

1 Introduction

1.1 Background

Geyer and Meeden (2005) introduced the idea of fuzzy (also called abstract randomized) hypothesis tests and confidence intervals. Their theory is a reinterpretation of the classical theory of uniformly most powerful (UMP) one-tailed tests and the uniformly most powerful unbiased (UMPU) two-tailed tests, which has been accepted as a key part of the theory of hypothesis tests since it was introduced by Neyman and Pearson in the 1930's.

Those hypothesis tests (or the confidence intervals that are dual to them) are randomized, that is, they involve additional artificial randomization independent of the randomness in the data. So two statisticians can analyze exactly the same data by exactly the same procedure, and arrive at opposite decisions because of the artificial randomization. Consequently, these procedures are not widely used in applied statistics.

Geyer and Meeden (2005) argue that this can be fixed by removing the artificial randomness. They bring in the distinction between a random variable, which is a theoretical construct described by a probability distribution, and a *realization* of the random variable, which is just a number purportedly obtained from some random process described by that probability distribution. Classical UMP and UMPU tests (and the confidence intervals dual to them) use realized artificial randomness. That is why different statisticians get different results for the same data and same procedure. Geyer and Meeden (2005) just say no. Leave the randomization abstract. Just describe the distribution of the random variable, and leave it at that. If users really want realizations, then they can simulate such themselves.

So Geyer and Meeden (2005) in no way disagree with the classical theory of UMP and UMPU tests. They only think we should leave the randomization abstract, so the results of any analysis are unique, and all analysts agree (on the abstract random variable described by a probability distribution that is the result).

1.2 Fuzzy Procedures

A classical randomized test rejects the null hypothesis with probability given by a *critical function* ϕ. When the observed data is x, then we reject the null with probability $\phi(x)$. But, of course, it also depends on the significance level α and the hypothesized value θ of the parameter under the null hypothesis. So we, following Geyer and Meeden (2005), write it $\phi(x, \alpha, \theta)$. By making the test *abstract randomized* Geyer and Meeden (2005) mean: describe this function ϕ and stop there.

They call these procedures *fuzzy* with the same meaning as in fuzzy set theory (Klir, et al., 1997). A set in ordinary mathematics can be described by its *indicator function* which is zero-or-one-valued. A *fuzzy* set in fuzzy set theory can be described by its *membership function* which takes values in the closed interval $[0, 1]$. Values strictly between zero and one correspond to points we are unsure whether or not they are in the fuzzy set and the value of the membership function says how much we think they are in or out.

Statisticians and others familiar with probability theory will note that probability is also given by numbers between zero and one, and probability also is thought to describe uncertainty, so is fuzzy set theory just probability theory under another name? No. The operations of fuzzy set theory (Klir, et al., 1997) have nothing to do with probability theory. It is really different. But Geyer and Meeden (2005) do not use any of those operations. They just use some of the terminology, mainly membership functions. They also use the term *crisp*.

A fuzzy set is *crisp* if its membership function is zero-or-one valued (so we are certain about which points are in the set and which are not, and this is just like the indicator function of a set in ordinary mathematics). So crisp is just the fuzzy set theory way of saying ordinary.

So classical confidence intervals (which are ordinary sets) are crisp fuzzy sets described by zero-or-one-valued membership functions. To have abstract randomized confidence intervals (dual to classical UMP and UMPU tests) Geyer and Meeden (2005) need general fuzzy sets described by membership functions that take values between zero and one.

Geyer and Meeden (2005) say the following about the critical function.

- The function $\phi(\,\cdot\,, \alpha, \theta)$ they call the fuzzy (or abstract randomized) decision function for the hypothesis test having significance level α and null hypothesis θ.

- The function $1 - \phi(x, \alpha, \,\cdot\,)$ they call the (membership function in the sense of fuzzy set theory) of the fuzzy confidence interval with coverage $1 - \alpha$ and observed data x.

- The function $\phi(x, \,\cdot\,, \theta)$ they call the (distribution function of) the abstract randomized (or fuzzy) P-value for the hypothesis test having null hypothesis θ and observed data x.

That the function $\phi(x, \,\cdot\,, \theta)$ is a distribution function is Theorem 1 below.

That the function $1 - \phi(x, \alpha, \cdot\,)$ is a membership function is obvious. If ϕ takes values in $[0, 1]$. then so does $1 - \phi$.

Geyer and Meeden (2005) only discuss fuzzy hypothesis tests and confidence intervals based on UMP and UMPU theory. Their discussants discuss other kinds. Papers citing Geyer and Meeden (2005) discuss other kinds. Every randomized test can be described by a critical function as described above, and then we have the interpretations listed above.

For UMP and UMPU fuzzy confidence intervals, we can also prove that they are convex in the sense of fuzzy set theory (Theorem 5 below).

1.3 More on Interpretation

For a one-tailed test, including UMP, the null and alternative hypotheses are

$$H_0 : \text{true unknown parameter} \le \theta \tag{1a}$$

$$H_1 : \text{true unknown parameter} > \theta \tag{1b}$$

for an upper-tailed test, and the same with the inequalities reversed for a lower-tailed test. The θ in these equations is the θ in the critical function $\phi(x, \alpha, \theta)$.

For a two-tailed test, including UMPU, the null and alternative hypotheses are

$$H_0 : \text{true unknown parameter} = \theta \tag{2a}$$

$$H_1 : \text{true unknown parameter} \ne \theta \tag{2b}$$

and the θ in these equations is the θ in the critical function $\phi(x, \alpha, \theta)$.

The power function of the test is

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

where θ' varies over the parameter space and θ is fixed at the value hypothesized under the null hypothesis.

- The *exactness* property is $p(\theta, \theta) = \alpha$ for all θ.
- The *unbiasedness* property is $p(\theta', \theta) \ge \alpha$, for all θ' and θ.
- The *UMP* property is $p(\theta', \theta) \ge \tilde{p}(\theta', \theta)$ whenever $\theta' < \theta$ for a lower-tailed test, whenever $\theta' > \theta$ for an upper-tailed test, and whenever $\theta' \ne \theta$ for a two-tailed test, and all power functions $\tilde{p}$ of other tests in the class of interest (satisfying exactness for UMP and satisfying exactness and unbiasedness for UMPU).

When the data are discrete, only a randomized test can have the exactness property (that's why randomized tests were invented by Neyman and Pearson).

The interpretation of the fuzzy P-value is simplest. The function $\phi(x, \cdot\,, \theta)$ is a distribution function. Let P denote a random variable that has this distribution. Then the test that rejects the null when $P \le \alpha$ is the classical UMP or

UMPU test. All Geyer and Meeden (2005) are saying in this case is report the *distribution of* P and stop there rather than going on to generating a number that is purported to be a realization of P or comparing that to α and making a decision. They claim that that distribution (the fuzzy P-value) is just as easy to interpret as a classical (crisp) P-value if one properly takes into account the equivocalness of P-values of intermediate size (near 0.05 by convention).

The value of thinking of a fuzzy confidence interval as a fuzzy set comes from the partial credit interpretation. As good frequentists, we know that the performance of a confidence interval is measured by averaging over all possible data. The exactness property is

$$E_\theta\{1 - \phi(X, \alpha, \theta)\} = 1 - \alpha, \qquad \text{for all } \theta.$$

in which we see that when $\phi(x, \alpha, \theta)$ is strictly between zero and one, the expectation gives partial credit for coverage.

Define $c(\theta', \theta) = 1 - p(\theta', \theta)$. Call it the coverage function for the fuzzy confidence interval. It is the probability that the interval covers θ when θ' is the true unknown parameter value. Then we have the same three properties. Here we only consider UMPU two-tailed intervals.

- The *exactness* property is $c(\theta, \theta) = 1 - \alpha$ for all θ. We have the exact desired coverage, unlike crisp confidence intervals.
- The *unbiasedness* property is $c(\theta', \theta) \leq 1 - \alpha$, for all θ' and θ. The fuzzy confidence interval has a higher probability of covering the true unknown parameter value than any other parameter value.
- The UMP property is $c(\theta', \theta) \leq \tilde{c}(\theta', \theta)$ for all θ' and θ where $\tilde{c}$ is the coverage function for any other fuzzy confidence interval satisfying exactness and unbiasedness. The fuzzy confidence interval dual to the UMPU test has less coverage for any false parameter value than any other fuzzy confidence interval having the exactness and unbiasednessi properties.

1.4 Computing

So everything depends on the critical function $\phi(x, \alpha, \theta)$. We want to provide computation of the critical function for some discrete exponential families that arise commonly in statistical inference (Section 1.5 below). See Section 1.8 below for more on UMP and UMPU procedures and critical functions. For more, see Geyer and Meeden (2005). For even more, see Lehmann (1959).

We want our implementation of the critical function to efficiently vectorize over a vector of α values for computation of fuzzy P-values and vectorize over a vector of θ values for computation of fuzzy confidence intervals. More about computation starting with Section 5 below.

1.5 Families of Distributions

The discrete exponential families we are interested in are

- the binomial distribution (more strictly, any binomial exponential family of distributions — different sample sizes determine different exponential families), needed not only for binomial data but also for comparison of the distributions of two independent Poisson random variables, two components of a multinomial random vector, or the UMP and UMPU competitors of McNemar's test (Sections 4.1, 4.4, 4.8, and 4.9 below),
- the Poisson distribution (more strictly, the Poisson exponential family of distributions), needed for Poisson data (Section 4.2 below),
- the negative binomial distribution (more strictly, any negative binomial exponential family of distributions — the shape parameter is considered known (otherwise we would not have an exponential family) so different shape parameters determine different exponential families), needed for negative binomial data (Section 4.3 below),
- Fisher's noncentral hypergeometric distribution (more strictly, any exponential family generated by a hypergeometric distribution — different numbers of population successes, population failures, and sample sizes determine different hypergeometric distributions, hence different exponential families), needed for comparison of the distributions of two independent binomial random variables or the UMP and UMPU competitors of Fisher's exact test (Sections 4.5 and 4.7 below), and
- any exponential family generated by a negative hypergeometric distribution, needed for comparison of the distributions of two independent negative binomial random variates (Section 4.6 below).

The first three of these are well-known and well supported in R with the usual quartet of d, p, q, and r functions (like `dbinom`, `pbinom`, `qbinom`, and `rbinom`). The last two of these do not have full R support (even in a CRAN package). The next to last does have some literature (Liao and Rosen, 2001) about algorithms for it and is implemented in R package `MCMCpack` (Martin, et al., 2022) but only d and r functions (`dnoncenhypergeom` and `rnoncenhypergeom`), which are insufficient for UMP and UMPU calculation (which need p and q functions too). The last does not even have that (neither literature nor R implementation).

1.6 More on Noncentral Hypergeometric

The hypergeometric distribution (Wikipedia contributors, 2023a) is the distribution of the number of successes in sampling without replacement from a finite population (that is, just like the binomial distribution except for sampling without replacement). The probability mass function is

$$f(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$

where there are K successes for N individuals in the population and x successes for n individuals in the sample. The support of this distribution is a bit tricky.

A binomial coefficient $\binom{m}{k}$ must satisfy $0 \leq k \leq m$, otherwise it is zero, and we have three binomial coefficients to deal with, so we must have

$$
\begin{gathered}
0 \leq x \leq K \\
0 \leq n - x \leq N - K \\
0 \leq n \leq N
\end{gathered}
$$

Hence in order for the distribution to exist we must have

$$
0 \leq n \leq N
$$

and in order for x to be in the support of the distribution we must have

$$
\max(0, n - (N - K)) \leq x \leq \min(K, n)
$$

The exponential family generated by this distribution, sometimes called Fisher's noncentral hypergeometric distribution (Fisher, 1935; Cornfield, 1956; Agresti, 1992; Liao and Rosen, 2001; Wikipedia contributors, 2021) has probability mass function

$$
f_\theta(x) = \frac{e^{x\theta} \binom{K}{x} \binom{N-K}{n-x}}{c(\theta)}
$$

where

$$
c(\theta) = \sum_{x = \max(0, n + K - N)}^{\min(K, n)} e^{x\theta} \binom{K}{x} \binom{N - K}{n - x}
$$

where x is the canonical statistic, θ is the canonical parameter, and c is the Laplace transform of the hypergeometric distribution.

1.7 More on Noncentral Negative Hypergeometric

The negative hypergeometric distribution (Wikipedia contributors, 2023b) is like the hypergeometric distribution except with inverse sampling. (Like the negative binomial distribution, the random variable of interest is the number of successes before the r-th failure, but like the hypergeometric distribution sampling is without replacement.) The probability mass function is

$$
f(x) = \frac{\binom{x+r-1}{x} \binom{N-r-x}{K-x}}{\binom{N}{K}}
$$

where there are K successes for N individuals in the population and x successes and r failures in the sample.

We have three binomial coefficients to deal with, so we must have

$$
\begin{gathered}
0 \leq x \leq x + r - 1 \\
0 \leq K - x \leq N - r - x \\
0 \leq K \leq N
\end{gathered}
$$

Hence in order for the distribution to exist we must have

$$0 \leq K \leq N \text{ and } 1 \leq r \leq N - K$$

and in order for x to be in the support of the distribution we must have

$$0 \leq x \leq K$$

The exponential family generated by this distribution has probability mass function

$$f_\theta(x) = \frac{e^{x\theta} \binom{x+r-1}{x} \binom{N-r-x}{K-x}}{c(\theta)} \tag{3}$$

where

$$c(\theta) = \sum_{x=0}^{K} e^{x\theta} \binom{x+r-1}{x} \binom{N-r-x}{K-x}$$

where x is the canonical statistic, θ is the canonical parameter, and c is the Laplace transform of the negative hypergeometric distribution.

1.8 UMP and UMPU

The following is mostly taken from Geyer and Meeden (2005).

1.8.1 Significance Level Zero or One

The only test that has significance level zero must reject the null hypothesis almost surely, hence for all possible values of a discrete test statistic. Thus we define

$$\phi(x, 0, \theta) = 0, \qquad \text{for all } x \text{ and } \theta. \tag{4}$$

Similarly

$$\phi(x, 1, \theta) = 1, \qquad \text{for all } x \text{ and } \theta. \tag{5}$$

In what follows, we need (4) because that will not follow from the equations we use when $\alpha > 0$ when the sample space is infinite (although it will be a limit of those equations). In contrast, we will not need (5) because that will be a special case of the equations we use when $\alpha > 0$.

1.8.2 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 and alternative hypotheses

$$\begin{aligned} H_0 &: \text{true unknown parameter} \leq \theta \\ H_1 &: \text{true unknown parameter} > \theta \end{aligned}$$

and critical function ϕ for size α defined by

$$\phi(x, \alpha, \theta) = \begin{cases} 1, & T(x) > C \\ \gamma, & T(x) = C \\ 0, & T(x) < C \end{cases} \tag{6}$$

where the constants γ and C are determined by

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

The description of the analogous lower-tailed test is the same except that all inequalities are reversed.

1.8.3 UMPU

Lehmann (1959, pp. 126–127) says for a one-parameter regular full exponential family model with canonical statistic $T(X)$ and canonical parameter θ there exists a UMPU test having null and alternative hypotheses

$$\begin{aligned} H_0 &: \text{true unknown parameter} = \theta \\ H_1 &: \text{true unknown parameter} \neq \theta \end{aligned}$$

and critical function ϕ for size α defined by

$$\phi(x, \alpha, \theta) = \begin{cases} 1, & T(x) < C_1 \\ \gamma_1, & T(x) = C_1 \\ 0, & C_1 < T(x) < C_2 \\ \gamma_2, & T(x) = C_2 \\ 1, & C_2 < T(x) \end{cases} \tag{7}$$

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

$$E_\theta\{\phi(X, \alpha, \theta)\} = \alpha \tag{8a}$$

$$E_\theta\{T(X)\phi(X, \alpha, \theta)\} = \alpha E_\theta\{T(X)\} \tag{8b}$$

That both sides of (8b) are finite is guaranteed by the regular full exponential family assumption.

1.8.4 Two-Point Sample Space

In case the sample space contains only two points, without loss of generality, we take $T(X)$ to have a Bernoulli distribution so $C_1 = 0$ and $C_2 = 1$. Let $p = \Pr\{T(X) = 1\}$. Then (8a) and (8b) become

$$\gamma_1(1 - p) + \gamma_2 p = \alpha \tag{9a}$$

$$\gamma_2 p = \alpha p \tag{9b}$$

which have solution $\gamma_1 = \gamma_2 = \alpha$. So in this case the UMPU test ignores both data and the hypothesized value p under the null hypothesis: the test rejects the null hypothesis with probability α regardless of either of them.

For any randomized test to be unbiased we must have

$$\gamma_1(1-p) + \gamma_2 p \geq \alpha$$

for all p and, together with (9a) holding when p is the value hypothesized under the null hypothesis, this clearly implies $\gamma_1 = \gamma_2 = \alpha$, so there is only one unbiased test, and it is, of course, UMP within the class consisting of that single test.

We only paid this case special attention because, if we don't know about it, then it seems odd.

2 Calculating UMP and UMPU

2.1 UMP

In (6) the constant C is clearly any $(1-\alpha)$-th quantile of the distribution of $T(X)$ for the parameter value θ. If the event $T(X) = C$ has probability zero, then the test is effectively not randomized and the value of γ is irrelevant (can be chosen arbitrarily). Otherwise

$$\gamma = \frac{\alpha - \Pr_\theta\{T(X) > C\}}{\Pr_\theta\{T(X) = C\}}. \tag{10}$$

By definition of $(1-\alpha)$-th quantile

$$\Pr_\theta\{T(X) > C\} \leq \alpha \leq \Pr_\theta\{T(X) \geq C\}$$

so (10) is always between zero and one (inclusive).

2.2 UMPU

In (7), if $C_1 = C_2 = C$, then $\gamma_1 = \gamma_2 = \gamma$ also. This occurs only in a very special case. Define

$$p = \Pr_\theta\{T(X) = C\} \tag{11a}$$
$$\mu = E_\theta\{T(X)\} \tag{11b}$$

Then in order to satisfy (8a) and (8b) we must have

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

which solved for γ and C gives

$$\gamma = 1 - \frac{1-\alpha}{p} \tag{12a}$$
$$C = \mu \tag{12b}$$

Thus this special case occurs only when $\Pr_\theta\{T(X) = \mu\}$ is nonzero, and then only for very large significance levels: $\alpha \geq 1 - p$. Hence this special case is of no practical importance, although it is of some computational importance to get every case right, no weird bogus results or crashes in unusual special cases.

Returning 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 C_1 and C_2 presently.) With μ still defined by (11b) and with the definitions

$$p_i = \Pr_\theta\{T(X) = C_i\}, \qquad i = 1, 2 \tag{13a}$$
$$P_1 = \Pr_\theta\{T(X) < C_1\} \tag{13b}$$
$$P_2 = \Pr_\theta\{T(X) > C_2\} \tag{13c}$$
$$M_1 = E_\theta\{T(X) I_{(-\infty, C_1)}[T(X)]\} \tag{13d}$$
$$M_2 = E_\theta\{T(X) I_{(C_2, \infty)}[T(X)]\} \tag{13e}$$

Then (8a) and (8b) become

$$P_1 + \gamma_1 p_1 + \gamma_2 p_2 + P_2 = \alpha \tag{14a}$$
$$M_1 + \gamma_1 C_1 p_1 + \gamma_2 C_2 p_2 + M_2 = \alpha\mu \tag{14b}$$

which solved for γ_1 and γ_2 give

$$\gamma_1 = \frac{\alpha(C_2 - \mu) + (M_1 - C_2 P_1) + (M_2 - C_2 P_2)}{p_1(C_2 - C_1)} \tag{15a}$$
$$\gamma_2 = \frac{\alpha(\mu - C_1) - (M_2 - C_1 P_2) - (M_1 - C_1 P_1)}{p_2(C_2 - C_1)} \tag{15b}$$

If (15a) and (15b) are both between 0 and 1 (inclusive), then C_1 and C_2 have been correctly determined. So this gives us (implicitly) an algorithm: keep trying different C_1 and C_2 until (15a) and (15b) both compute numbers between 0 and 1 (inclusive). More sophisticated algorithms will be developed below.

2.3 Alternative Formulas for UMPU

Geyer and Meeden (2005) give alternative formulas for γ_1 and γ_2. Define

$$P_{12} = \Pr_\theta\{C_1 < T(X) < C_2\} \tag{16a}$$
$$M_{12} = E_\theta\{T(X) I_{(C_1, C_2)}[T(X)]\} \tag{16b}$$

Then

$$1 - \gamma_1 = \frac{(1 - \alpha)(C_2 - \mu) + M_{12} - C_2 P_{12}}{p_1(C_2 - C_1)} \tag{17a}$$
$$1 - \gamma_2 = \frac{(1 - \alpha)(\mu - C_1) - M_{12} + C_1 P_{12}}{p_2(C_2 - C_1)} \tag{17b}$$

where p_1 and p_2 are still given by (13a).

These equations are useful in some theoretical contexts, especially in the case where C_1 and C_2 are adjacent atoms so $M_{12} = P_{12} = 0$.

It might be thought that these equations are also computationally useful, avoiding catastrophic cancellation when their values are near zero. But catastrophic cancellation is only avoided if M_{12} and P_{12} can be computed without catastrophic cancellation, which the naive method (subtraction of cumulative distribution function values) fails to do.

Subtracting μ times (32a) from (32b) gives

$$E_\theta\{(X - \mu)\phi(X, \alpha, \theta)\} = 0$$

and this can be rewritten

$$E_\theta\{(\mu - X)I_{(-\infty,\mu)}(X)\phi(X, \alpha, \theta)\} = E_\theta\{(X - \mu)I_{(\mu,\infty)}(X)\phi(X, \alpha, \theta)\} \quad (18)$$

and this can be used together with either of (32a) or (32b) to determine the UMPU critical function ϕ.

Since the expectations in (18) are not easy to compute, (18) may not be computationally helpful. It does show that UMPU tests are equal tailed, not in the usual sense of equal probabilities in each tail but rather in the sense of equal contributions (with opposite sign) to the expectation of $X - \mu$ in each tail.

3 Theorems about UMP and UMPU

3.1 Fuzzy *P*-Values

3.1.1 Continuity, Monotonicity, Piecewise Linearity

A random variable or its distribution is discrete in the sense of probability theory if it has countable support, in which case the smallest support (event having probability one) is the set of *atoms* (points having positive probability), and we will call that *the support*.

We will say that a (probabilistically) discrete random variable or its distribution is *order discrete* if the support, considered as a subspace of the real line, is a discrete ordered set: for every point x of the support except the least (if there is a least) there is a next lower point in the support, and for every point x of the support except the greatest (if there is a greatest) there is a next higher point in the support. This property holds for all of our distributions of interest (which are integer valued).

It holds for all discrete distributions of applied statistics as far as I know. It is easy enough to construct a probabilistically discrete but not order discrete example. A mixture of an atom at -1 and $1/X$ where X is Poisson does the job. No point above -1 is next to it. But I do not know what an example of such a distribution would be that has a real application.

Theorem 1. *Assume a regular full one-parameter exponential family such that the canonical statistic is probabilistically and order discrete. For either UMP or UMPU the function* $F_{x,\theta} : \alpha \mapsto \phi(x, \alpha, \theta)$ *is continuous, piecewise linear, nondecreasing on the interval* $[0, 1]$*, and maps onto the interval* $[0, 1]$*.* $F_{x,\theta}$ *is strictly increasing near* α *such that* $0 < F_{x,\theta}(\alpha) < 1$*. Hence* $F_{x,\theta}$ *is the distribution function of a continuous random variable having support that is a subinterval of* $[0, 1]$*.*

For UMP the distribution having this distribution function is uniform on an interval.

In aid of proving this theorem we state an algorithm (Algorithm 1 below).

Algorithm 1 Computing the UMPU Critical Function

Assume a regular full one-parameter exponential family such that the canonical statistic is order discrete. For fixed θ, the following computes the critical function of the UMPU test $\phi(x, \alpha, \theta)$ for all x and $\alpha > 0$.

1. Start with $\alpha = 1$.
   - (a) If μ given by (11b) is an atom, then $\phi(x, \alpha, \theta)$ is given by (7) with $C_1 = C_2 = \mu$ and $\gamma_1 = \gamma_2 = \gamma$ given by (11a), (12a), and (12b) over the range of α such that (12a) is between zero and one.
   - (b) If μ given by (11b) 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 (7) with γ_1 and γ_2 given by (13a), (13b), (13c), (13d), (13e), (15a), and (15b) over the range of α such that both (15a) and (15b) are between zero and one.
2. Start with the lowest α for which $\phi(x, \alpha, \theta)$ was determined in step 1 or a previous iteration of step 2. At this point, either γ_1 or γ_2 is zero (or both are).
   - (a) If γ_1 is zero, then decrease C_1 to the next point of the sample space below the current value and set $\gamma_1 = 1$.
   - (b) If γ_2 is zero, then increase C_2 to next point of the sample space above the current value and set $\gamma_2 = 1$.
   - (c) Now $\phi(x, \alpha, \theta)$ is given by (12a) with γ_1 and γ_2 given by (13a), (13b), (13c), (13d), (13e), (15a), and (15b) over the range of α such that both (15a) and (15b) are between zero and one.
3. Repeat step 2 until the whole range $0 < \alpha \leq 1$ is covered.

Without the order discrete assumption, it could be that there would be no next point to select in step 2a or 2b of the algorithm (or even points C_1 and C_2 to choose in 1b).

In case the sample space is infinite, which among our exponential families of interest is for Poisson and negative binomial, step 2 of the algorithm may need to be repeated an infinite number of times, as it is clear (from the algorithm and the proof of the theorem) that C_1 visits every point of the sample space between μ and the lower bound (which could be $-\infty$ but is finite for all our exponential families of interest) and C_2 visits every point of the sample space between μ and the upper bound (which is ∞ for the Poisson and negative binomial families).

Thus, strictly speaking, if one considers termination after a finite number of steps to be part of the definition of algorithm, this is not an algorithm. But we will not worry about that. Just consider it a theoretical construction rather than a computer algorithm. (More on this after the proof.)

Proof of Theorem 1. In the UMP case (10) computes a number strictly between 0 and 1 if and only if α is strictly between $\Pr_\theta\{T(X) > C\}$ and $\Pr_\theta\{T(X) \geq C\}$. On this interval, γ is a linear function of α and goes from 0 to 1. Thus the fuzzy P-value has the continuous uniform distribution on the interval with endpoints $\Pr_\theta\{T(X) > C\}$ and $\Pr_\theta\{T(X) \geq C\}$, where C is the observed value of $T(X)$.

For UMPU we apply Algorithm 1. What it computes is continuous in α because steps 2a and 2b do not change the critical function (merely its description) and the other steps change the critical function continuously in α using (12a) in step 1a or (15a) and (15b) in step 1b or 2c. Moreover, (12a), (15a), and (15b) are linear and strictly increasing in α. So this proves $F_{x,\theta}$ is continuous, piecewise linear, and increasing on the open interval where its values are strictly between 0 and 1. And we have proved that the supremum of its values is 1 but still have to prove the infimum is 0.

If we ever have $C_1 < T(x) < C_2$ at any point in execution of Algorithm 1, then we have $F_{x,\theta}(\alpha) = 0$ for the rest of the execution of the algorithm. Thus the only way we can have $F_{x,\theta}(\alpha) > 0$ for the whole execution is, if from some point in the execution onward either step 2a or step 2b is never executed again (hence the other is executed infinitely often). This can only happen with an infinite sample space. With the families of interest, all of which are bounded below, this would mean that C_1 is eventually constant for an infinite number of iterations of the algorithm while $C_2 \to \infty$. We will do this case; the other case is similar (just swap left and right on the number line).

By dominated convergence, $C_2 \to \infty$ implies $P_2 \to 0$ and $M_2 \to 0$ (for a dominating function we can take $|T(X)|$, which is guaranteed to be integrable by the model being a regular full exponential family. Plugging those limits into (14a) and (14b) gives

$$P_1 + \gamma_1 p_1 = \alpha \tag{19a}$$

$$M_1 + \gamma_1 C_1 p_1 = \alpha\mu \tag{19b}$$

because $0 \leq p_2 \leq p_2 + P_2 \to 0$ and eventually (when C_2 is positive) $0 \leq C_2 p_2 \leq C_2 p_2 + M_2 \to 0$ and $0 \leq \gamma_2 \leq 1$. But (19a) and (19b) are the equations for a lower-tailed UMP test, so the solutions are given by (10) with the inequality reversed

$$\gamma_2 = \frac{\alpha - P_1}{p_1}.$$

Hence for sufficiently small α this will go negative unless $P_1 = 0$ which implies that C_1 is the lower bound of the sample space, in which case α can decrease to zero. Thus Algorithm 1 can make α arbitrarily close to zero, and when it does $F_{x,\theta}(\alpha)$ also gets arbitrarily close to zero. This proves the assertions about continuity on $[0, 1]$. □

As noted before the proof of Theorem 1, Algorithm 1 is not, strictly speaking, an algorithm because it can need to do an infinite amount of work. However, if we are only interested in one fixed x rather than all x, the proof of Theorem 1 shows that we do an infinite amount of work only if we are doing a UMPU two-tailed test and $T(x)$ is a boundary point of its sample space. In that case we will have to use a convergence tolerance and stop when α gets close enough to zero in order to have a computer algorithm. (Theorem 2 below helps in choosing the convergence tolerance.)

However, Algorithm 1 still does more work than necessary.

3.1.2 Endpoint Behavior

The behavior of the (distribution function of the) fuzzy P-value where its values are near 1 has already been described (Step 1 of Algorithm 1). In either case (Step 1a or 1b), it is piecewise linear, continuous, and satisfies (5).

For any continuous function, we say a point x in its domain is a *knot* if the first derivative or some higher derivative is discontinuous at x. If the function is given by a formula, the formula is different on each side of the knot. The way R draws graphs of functions (connecting the dots) these discontinuities are smoothed out unless the values at the knots are included in the input to the plot.

Theorem 2. *With the assumptions of Theorem 1, in case the canonical statistic $T(x)$ is on the boundary of the sample space of $T(X)$, which is unbounded (in the other direction),*

$$\frac{\phi(x, \alpha, \theta)}{\alpha} \to \frac{1}{\Pr_\theta\{T(X) = T(x)\}}, \qquad \text{as } \alpha \to 0.$$

In all other cases $\phi(x, \cdot\,, \theta)$ is piecewise linear on $[0, 1]$ with a finite number of knots, continuous, and satisfies (4).

All of this was proved in the proof of Theorem 1.

3.2 Fuzzy Confidence Intervals

3.2.1 Some Fundamentals of Exponential Families

The probability mass function of the canonical statistic of a discrete exponential family can be written

$$f_\theta(x) = e^{x\theta - c(\theta)} \lambda(x) \tag{20}$$

where θ is the canonical parameter, c is the cumulant function, and λ is positive when x is an atom and zero otherwise. For general (not necessarily discrete) families, the formula is similar (Geyer, 1990, Section 1.2).

The function c is called the cumulant function of the family. From probabilities summing to one we get

$$c(\theta) = \log \left(\sum_{x \in \mathbb{R}} e^{x\theta} \lambda(x) \right) \tag{21}$$

where the only terms in the sum that contribute are x such that $\lambda(x) > 0$, the atoms of the distribution. We take (21) to define the cumulant function on the whole real line, writing $c(\theta) = \infty$ if the sum does not exist. Then θ is the canonical parameter of the exponential family, and

$$\Theta = \{\, \theta \in \mathbb{R} : c(\theta) < \infty \,\} \tag{22}$$

is the canonical parameter space of the *full* exponential family containing the originally given exponential family (with the probability mass functions of the family being given by (20)). Cumulant functions are convex (Barndorff-Nielsen, 1978, Theorem 7.1). So the full canonical parameter space (22) is an interval of real numbers. It need not be the whole real line. For negative binomial (22) is the half-line $(-\infty, 0)$.

The full family is *regular* if (22) is an open subset of the real numbers.

Derivatives of the cumulant function are cumulants (hence the name). The first two are

$$c'(\theta) = E_\theta(X) \tag{23a}$$

$$c''(\theta) = \operatorname{var}_\theta(X) \tag{23b}$$

(Barndorff-Nielsen, 1978, Theorem 8.1). These equations hold at all θ in the interior of (22), hence, for a *regular* full exponential family, for all $\theta \in \Theta$.

From (23b) we see that, unless the support is a single point, $c''(\theta) > 0$, so c is a strictly convex function (Barndorff-Nielsen, 1978, Theorem 7.1). Hence c' is a strictly increasing on Θ (assuming regular). Hence by the inverse function theorem, c' is an invertible function. So

$$\mu = c'(\theta) = E_\theta(X)$$

is a one-to-one mapping between the canonical parameter θ and the mean value parameter μ. And by moment generating function theory and the inverse function theorem, the mappings both ways are infinitely differentiable (a fact we will not use in this document). Here we only need that the mappings are both continuous (which is implied by differentiability), so both map an open interval to an open interval.

It is not obvious from what we have said so far, but the mean value parameter space (the range of the function $c' : \Theta \to \mathbb{R}$) is the interior of the convex hull of the support of the canonical statistic (Barndorff-Nielsen, 1978, Theorems 8.1 and 9.2).

Cumulant functions are lower semicontinuous on $\mathbb{R}$ (Barndorff-Nielsen, 1978, Theorem 7.1). So $c(\theta) = \infty$ implies

$$c(\theta_n) \to \infty, \qquad \text{as } \theta_n \to \theta.$$

3.2.2 Convexity and Continuity

Lemma 3. *In (15a) and (15b), if C_1 and C_2 are chosen so that $C_1 \neq C_2$ and both γ_1 and γ_2 are strictly between zero and one (so these formulas define the critical function of a UMPU test), and if the support of the canonical statistic has more than two points (excluding the case described in Section 1.8.4 above), then $\partial \gamma_1 / \partial \theta > 0$ and $\partial \gamma_2 / \partial \theta < 0$.*

Proof. Under the assumptions, we have $C_1 < \mu < C_2$ by Theorem 1. To simplify notation assume $T(X) = X$ so we have a standard exponential family, define $j = 3 - i$, and define

$$\begin{aligned} A_{\text{in}} &= \{\, x : C_1 < x < C_2 \,\} \\ A_{\text{out}} &= \{\, x : x < C_1 \text{ or } C_2 < x \,\} \end{aligned}$$

(either of these can be empty, but not both by the more than two points assumption).

Now we can rewrite both (15a) and (15b) as

$$\begin{aligned} \gamma_i &= \frac{\alpha(C_j - \mu) + (M_i - C_j P_i) + (M_j - C_j P_j)}{p_i(C_j - C_i)} \\ &= \frac{E_\theta\{(X - C_j)[I_{A_{\text{out}}}(X) - \alpha]\}}{f_\theta(C_i)(C_j - C_i)} \\ &= \int \frac{(x - C_j)[I_{A_{\text{out}}}(x) - \alpha]}{C_j - C_i} \cdot \frac{f_\theta(x)}{f_\theta(C_i)} \lambda(x) \\ &= \int \frac{(x - C_j)[I_{A_{\text{out}}}(x) - \alpha]}{C_j - C_i} \cdot e^{(x - C_i)\theta} \lambda(x) \end{aligned}$$

where θ is now the canonical parameter, f_θ is the probability mass function, and λ is a discrete positive measure that does not depend on θ (Section 3.2.1 above).

This equation can be differentiated under the integral sign giving

$$\frac{\partial \gamma_i}{\partial \theta} = \int \frac{(x - C_i)(x - C_j)[I_{A_{\text{out}}}(x) - \alpha]}{C_j - C_i} \cdot e^{(x - C_i)\theta} \lambda(x) \tag{24}$$

(Ferguson, 1996, Lemma of Chapter 18). The dominating function in Ferguson's lemma can be taken to be a constant plus $e^{x(\theta - \varepsilon)} + e^{x(\theta + \varepsilon)}$ which is guaranteed to have finite integral for some $\varepsilon > 0$ by the assumption that we have a regular exponential family.

In case $i = 1$ and $0 < \alpha < 1$ in (24) the integrand is positive for $x \in A_{\text{out}} \cup A_{\text{in}}$ and zero otherwise. Hence the derivative is positive.

In case $i = 2$ and $0 < \alpha < 1$ in (24) the integrand is negative for $x \in A_{\text{out}} \cup A_{\text{in}}$ and zero otherwise. Hence the derivative is negative. ☐

Lemma 4. *In (10), if C is chosen so that γ is strictly between zero and one (so this formula defines the critical function of a UMP upper-tailed test), and if the support of the canonical statistic has more than one point, then $\partial\gamma/\partial\theta < 0$.*

For the UMP lower-tailed test all inequalities are reversed, so $\partial\gamma/\partial\theta > 0$.

Proof. As in the preceding lemma, we rewrite (10)

$$\begin{aligned}
\gamma &= \frac{E_\theta\{\alpha - I_{(C,\infty)}(X)\}}{f_\theta(C)} \\
&= \int [\alpha - I_{(C,\infty)}(x)] \cdot \frac{f_\theta(x)}{f_\theta(C)} \\
&= \int [\alpha - I_{(C,\infty)}(x)] e^{(x-C)\theta} \lambda(x)
\end{aligned}$$

(with the same notation as in the preceding lemma). This equation can be differentiated under the integral sign giving

$$\frac{\partial\gamma}{\partial\theta} = \int [\alpha - I_{(C,\infty)}(x)](x - C) e^{(x-C)\theta} \lambda(x) \tag{25}$$

(with argument about differentiation under the integral sign as in the preceding lemma). The integrand is negative when $x \neq C$ and zero otherwise. Hence the derivative is negative. ☐

Theorem 5. *With the assumptions of Theorem 1, a fuzzy confidence interval for the canonical parameter or any increasing function of it (including the mean value parameter) corresponding to a UMP or UMPU test for a regular full exponential family whose canonical statistic is probabilistically and order discrete is convex, and its membership function is continuous.*

Let μ denote the mean value parameter and x the observed value of the canonical statistic. The membership function is nondecreasing for $\mu < x$ and nonincreasing for $\mu > x$. If the maximum less than one, then the maximum occurs at $\mu = x$ where (12a) and (12b) and (15a) and (15b) all hold simultaneously. If the maximum of the membership function is one, then $\mu = x$ is one point where it is one.

To say a fuzzy set is convex is to say that each of the level sets $\{ x : I(x) \geq \gamma \}$ of its membership function I are convex (Geyer and Meeden, 2005, Section 1.3).

Note that this does not mean that the graph of the membership function or any part of it is convex or concave. See examples in Appendix B.

Proof. If the membership function of the fuzzy confidence interval for the canonical parameter θ is convex, then so is the same for any increasing function of θ.

The mean value parameter is an increasing function of the canonical parameter (Section 3.2.1 above).

From the lemmas, the negative derivatives are for γ_2 for the UMPU test or for γ of the UMP upper-tailed test. Thus these are decreasing functions of θ. If they ever decrease to zero, then we can change the representation of the critical function without changing its value, by increasing C_2 or C to the next possible value and setting γ_2 or γ to one.

And similarly for γ_1 for the UMPU test or for γ for the UMP lower-tailed test. These are increasing functions of θ. If they ever increase to one, then we can change the representation of the critical function without changing its value, by increasing C_1 or C to the next possible value and setting γ_1 or γ to zero.

Thus the critical function $\phi(x, \alpha, \theta)$ is continuous in θ, and C_1 and C_2 and C can only increase as θ increases. Hence the membership function of the fuzzy confidence interval corresponding to the upper-tailed UMP test is nondecreasing in θ, for the lower-tailed UMP test it is nonincreasing, and for the two-tailed UMPU test it may be constant at zero for a while (when $C_2 < x$), then increase (when $C_2 = x$), then be constant at one for a while (when $C_1 < x < C_2$), then decrease (when $C_1 = x$), then be constant at zero for a while (when $x < C_1$). But some of these parts may be omitted for certain x (more on this below). For any of these the fuzzy confidence interval is convex.

The only way we can have $\phi(x, \alpha, \theta)$ strictly between zero and one is when $x = C_1$ or $x = C_2$. If the latter, then $\mu \le C_2 = x$, and the membership function is decreasing by Lemma 3. If the former, then $x = C_1 \le \mu$, and the membership function is increasing by the same lemma. And, of course, the membership function is constant on intervals where it is equal to zero or one. This proves the last paragraph of the theorem statement.

So the proof is finished except that we still must look at the case where the mean value parameter μ crosses a possible value of the canonical statistic. There we may have (12a) and (12b) holding when $\mu = C$, but have (15a) and (15b) on either side. So we need to check that we have continuity here. Rather than (15a) and (15b), it is more convenient to use (17a) and (17b), which in this case (where C_1 and C_2 are adjacent and μ is between) become

$$1 - \gamma_1 = \frac{(1 - \alpha)(C_2 - \mu)}{f_\mu(C_1)(C_2 - C_1)}$$

$$1 - \gamma_2 = \frac{(1 - \alpha)(\mu - C_1)}{f_\mu(C_2)(C_2 - C_1)}$$

And these give us $1 - \gamma_1 \to (1 - \alpha)/f_\mu(C_1)$ agreeing with (12a) and $\gamma_2 \to 0$ as $\mu \to C_1$. The argument for $\mu \to C_2$ is similar. Thus we do have continuity when we switch from (12a) and (12b) to (15a) and (15b). $\square$

3.2.3 Endpoint Behavior

The UMPU test makes no sense when the null hypothesis is on the boundary of the mean value parameter space. Why do a two-tailed test when the null hypothesis says only one tail is possible?

But equations (7), (8a), and (8b) still make sense and define a test, which is the limit of tests done for all nearby parameter values hypothesized under the null hypothesis. 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).

Theorem 6. *With the assumptions of Theorem 1, if the support of $T(X)$ has a lower bound and L is the two-point set consisting of the two lowest atoms of the support, then*

$$1 - \phi(x, \alpha, \theta) \to (1 - \alpha) I_L\{T(x)\}, \qquad \text{as } \theta \to -\infty, \tag{26}$$

where θ is the canonical parameter. Similarly, if the support has an upper bound and L consists of the two highest atoms, then (26) holds with $-\infty$ replaced by $+\infty$.

Proof. We do the case where the support has a lower bound. The upper bound case is entirely analogous.

Without loss of generality, we assume $T(X) = X$, and use most of Section 3.2.1 above. Because X is bounded below (we assume), the full canonical parameter space (22) extends to $-\infty$. Because we assume the support of the distribution has more than one point (23b) cannot be zero, and μ and θ are strictly increasing continuous functions of each other.

Let $L = \{s_0, s_1\}$ with $s_0 < s_1$ be the set containing the two lowest points of the support of X. Let S denote the support. Since

$$\frac{f_\theta(s)}{f_\theta(s_0)} = e^{(s - s_0)\theta} \frac{\lambda(s)}{\lambda(s_0)}$$

the distribution converges to the distribution concentrated at s_0 as $\theta \to -\infty$ by monotone convergence.

Now

$$\frac{\mu(\theta) - s_0}{f_\theta(s_1)} = \sum_{s \in S \setminus \{s_0\}} (s - s_0) e^{(s - s_1)\theta} \frac{\lambda(s)}{\lambda(s_1)}$$

goes to $s_1 - s_0$ by monotone convergence as $\theta \to -\infty$. Hence $\mu(\theta) \to s_0$ as $\theta \to -\infty$.

Now

$$\frac{\Pr_\theta\{T(X) > s_1\}}{f_\theta(s_1)} = \sum_{s \in S \setminus L} e^{(s - s_1)\theta} \frac{\lambda(s)}{\lambda(s_1)}$$

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

And these facts together imply

$$\begin{aligned} \Pr_\mu\{T(X) = s_0\} &= \frac{s_1 - \mu}{s_1 - s_0} + o(\mu - s_0) \\ \Pr_\mu\{T(X) = s_1\} &= \frac{\mu - s_0}{s_1 - s_0} + o(\mu - s_0) \\ \Pr_\mu\{T(X) > s_1\} &= o(\mu - s_0) \end{aligned} \tag{27}$$

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

Now we claim that for low enough values of θ the UMPU test is given by (7) with $C_1 = s_0$ and $C_2 = s_1$ and γ_1 and γ_2 given by (17a) and (17b), which in this case become

$$\gamma_1 = 1 - \frac{(1-\alpha)(s_1 - \mu)}{\Pr_\mu\{T(X) = s_0\}(s_1 - s_0)}$$
$$\gamma_2 = 1 - \frac{(1-\alpha)(\mu - s_0)}{\Pr_\mu\{T(X) = s_1\}(s_1 - s_0)}$$

and clearly both converge to α as $\mu \to s_0$ hence both are between zero and one for low enough θ or μ and hence define the UMPU test. $\square$

This explains the behavior of the fuzzy confidence intervals for the binomial distribution for the two x values nearest each boundary in Figure 2 of Geyer and Meeden (2005). 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.

Theorem 7. *With the assumptions of Theorem 1, if the support of $T(X)$ has no upper bound, then C_1 and C_2 in (7) both go to infinity as the mean value parameter goes to infinity.*

Proof. Since $C_1 \leq \mu \leq C_2$, only the behavior of C_1 is at issue. And C_1 is a nondecreasing function of θ by the proof of Theorem 5. Hence either $C_1 \to \infty$ as $\mu \to \infty$ or there is a finite b such that $C_1 \leq b$ for all μ. We start a proof by contradiction by assuming the latter.

The full canonical parameter space (22) is an open interval (the regular full exponential family assumption). Let B denote its upper bound, which may be finite or infinite. We have two different arguments depending on whether B is finite or not.

First assume B is infinite, if $x_1 < x_2$, then

$$\frac{f_\theta(x_1)}{f_\theta(x_2)} = e^{(x_1 - x_2)\theta}.$$

As $\theta \to \infty$ the right-hand side goes to zero. But (for discrete data) probabilities are bounded by one so

$$f_\theta(x) \to 0, \qquad \text{as } \theta \to B \text{ and } \mu \to \infty. \tag{28}$$

Second assume B is finite, then

$$f_\theta(x) \leq e^{xB - c(\theta)}$$

but $c(\theta) \to \infty$ as $\theta \to B$ (end of Section 3.2.1 above). So again we have (28).

Now let A be any event that is bounded above (in the mean value parameter space, where X also takes values). Then for any nonnegative function g

$$E_\mu\{I_A(X)g(X)\} \to 0, \qquad \text{as } \mu \to \infty$$

by monotone convergence (because the integrand is decreasing after μ is greater than any element of A), provided the expectation exists for some μ greater than any element of A.

It follows that as $\mu \to \infty$

$$\begin{aligned} E_\mu\{I_{(-\infty,\mu)}(X)\phi(X,\alpha,\mu)\} &\to 0 \\ E_\mu\{XI_{(-\infty,\mu)}(X)\phi(X,\alpha,\mu)\} &\to 0 \end{aligned}$$

REVISED DOWN TO HERE

hence from (8a) and (8b)

$$\begin{aligned} E_\mu\{I_{(\mu,\infty)}(X)\phi(X,\alpha,\mu)\} &\to \alpha \\ E_\mu\{XI_{(\mu,\infty)}(X)\phi(X,\alpha,\mu)\}/\mu &\to \alpha \end{aligned}$$

so

$$E_\mu\left\{\left(\frac{X}{\mu} - 1\right) I_{(\mu,\infty)}(X)\phi(X,\alpha,\mu)\right\} \to 0$$

□

3.2.4 Summary

Fuzzy confidence intervals (more precisely, the membership function thereof, but we won't be pedantic about that here) behave as follows.

- If the support of the canonical statistic has two points, then the fuzzy confidence interval is constant, equal to $1-\alpha$ for all values of the parameter. Henceforth we assume the sample space has more than two points.

- If the observed value x of the canonical statistic is as low as possible, then the fuzzy confidence interval starts at $1-\alpha$ (when μ is as low as possible) and decreases to zero (at some point where μ is not as high as possible). And is zero thereafter (by convexity).

- If the observed value x of the canonical statistic is the next to lowest possible, then the fuzzy confidence interval starts at $1-\alpha$ (when μ is as small as possible) and increases. Depending on the value of α it may go all the way up to one, or may not. If it is less than one for all values of the parameter, then it reaches its maximum when $x = \mu$ and equations (12a) and (12b) hold. Unless x is one of the two highest points of the support, the fuzzy confidence interval decreases to zero and is zero thereafter.

- If the observed value x of the canonical statistic is not either of the two lowest possible values or either of the two highest possible values, then the fuzzy confidence interval starts at zero (when μ is as small as possible), stays at zero for a while, and then increases (as μ increases). Depending on the value of α it may go all the way up to one, or may not. If it is less than one for all values of the parameter, then it reaches its maximum when $x = \mu$ and equations (12a) and (12b) hold. After reaching the maximum, it decreases to zero, and is zero thereafter.

Appendix B has concrete examples of this behavior.

3.3 Models With Nuisance Parameters

UMP and UMPU theory extends to multiparameter exponential families when the parameter of interest θ is one of the canonical parameters (Lehmann, TSH, 1st ed., pp. 134–136).

Suppose the family has densities of the form

$$\frac{1}{c(\theta, \boldsymbol{\eta})} \exp\left(\theta T(x) + \sum_{i=1}^{k} \eta_i U_i(x)\right)$$

with respect to some measure on the sample space. Then the situation is exactly the same as described above except that the reference distribution of the test is the conditional distribution of $T(X)$ given $\mathbf{U}(X)$, which (a standard fact about exponential families) depends only on θ and not on the nuisance parameter $\boldsymbol{\eta}$.

3.3.1 UMP Tests With Nuisance Parameters

Now there exists a UMP test having null hypothesis $H_0 = \{\, \vartheta : \vartheta \le \theta \,\}$, alternative hypothesis $H_1 = \{\, \vartheta : \vartheta > \theta \,\}$, and significance level α, and its critical function ϕ is defined by

$$\phi(x, \alpha, \theta) = \begin{cases} 1, & T(x) > C[\mathbf{U}(x)] \\ \gamma[\mathbf{U}(x)], & T(x) = C[\mathbf{U}(x)] \\ 0, & T(x) < C[\mathbf{U}(x)] \end{cases} \tag{29}$$

where the functions γ and C are determined by

$$E_\theta\{\phi(X, \alpha, \theta) \mid \mathbf{U}(X)\} = \alpha. \tag{30}$$

Everything is exactly the same as for the one-parameter case except for the conditioning on $\mathbf{U}(x)$. The only point of the discussion is that the test is UMP whether considered conditionally or unconditionally.

As before, the UMP upper-tailed test is obtained by reversing all the inequalities above.

3.3.2 UMPU Tests With Nuisance Parameters

Now there exists a UMPU test having null hypothesis $H_0 = \{\, \vartheta : \vartheta = \theta \,\}$, alternative hypothesis $H_1 = \{\, \vartheta : \vartheta \neq \theta \,\}$, and significance level α, and its critical function ϕ is defined by

$$\phi(x, \alpha, \theta) = \begin{cases} 1, & T(x) < C_1[\mathbf{U}(x)] \\ \gamma_1[\mathbf{U}(x)], & T(x) = C_1[\mathbf{U}(x)] \\ 0, & C_1[\mathbf{U}(x)] < T(x) < C_2[\mathbf{U}(x)] \\ \gamma_2[\mathbf{U}(x)], & T(x) = C_2[\mathbf{U}(x)] \\ 1, & C_2[\mathbf{U}(x)] < T(x) \end{cases} \tag{31}$$

where the functions γ_1, γ_2, C_1, and C_2 are determined by

$$E_\theta\{\phi(X, \alpha, \theta) \mid \mathbf{U}(X)\} = \alpha \tag{32a}$$

$$E_\theta\{T(X)\phi(X, \alpha, \theta) \mid \mathbf{U}(X)\} = \alpha E_\theta\{T(X) \mid \mathbf{U}(X)\} \tag{32b}$$

Again, the point is that the test is UMPU whether considered conditionally or unconditionally.

4 Calculations For Distributions

4.1 Binomial

Let $X \sim \text{Bin}(n, p)$ with $0 < p < 1$.

All of the quantities in (15a) and (15b) are easily calculated (in R) except possibly M_1 and M_2. Actually, as Lehmann points out (TSH, 1st, ed., pp. 128–129), these are also easy to calculate

$$\begin{aligned} M_1 &= \sum_{x=0}^{C_1 - 1} x \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=1}^{C_1 - 1} x \binom{n}{x} p^x (1-p)^{n-x} \\ &= np \sum_{x=1}^{C_1 - 1} \binom{n-1}{x-1} p^{x-1} (1-p)^{n-x} \\ &= np \sum_{y=0}^{C_1 - 2} \binom{n-1}{y} p^y (1-p)^{n-1-y} \end{aligned}$$

and the last sum is just a binomial probability for the $\text{Bin}(n-1, p)$ distribution, that is, M_1 is calculated in R (with the obvious definitions of the variables) by

```
n * p * pbinom(c1 - 2, n - 1, p)
```

Let's check that this does not crash for small values of C_1.

```
n <- 10
p <- 2 / 3
c1 <- seq(0, n * p)
c1

## [1] 0 1 2 3 4 5 6

m1 <- n * p * pbinom(c1 - 2, n - 1, p)
m1

## [1] 0.0000000000 0.0000000000 0.0003387018 0.0064353334 0.0552083863
## [6] 0.2828159664 0.9656387068
```

And compare with obvious calculation

```
m1.too <- cumsum((c1 - 1) * dbinom(c1 - 1, n, p))
all.equal(m1, m1.too)

## [1] TRUE
```

Similarly, exchanging successes and failures,

$$M_2 = np \sum_{y=C_2}^{n-1} \binom{n-1}{y-1} p^y (1-p)^{n-1-y}$$

So M_2 is calculated by

`n * p * pbinom(c2 - 1, n - 1, p, lower.tail = FALSE)`

Try that out too

```
n <- 10
p <- 2 / 3
c2 <- seq(n, n * p) |> rev()
c2

## [1]  7  8  9 10

m2 <- n * p * pbinom(c2 - 1, n - 1, p, lower.tail = FALSE)
m2

## [1] 2.5145218 0.9537841 0.1734153 0.0000000
```

And compare with obvious calculation

```
m2.too <- rev(cumsum(rev((c2 + 1) * dbinom(c2 + 1, n, p))))
all.equal(m2, m2.too)

## [1] TRUE
```

4.2 Poisson

Let $X \sim \text{Pois}(\mu)$ with $0 < \mu$.

Again, all of the quantities in (15a) and (15b) are easily calculated except possibly M_1 and M_2. Do they work like the binomial case? Yes!

$$
\begin{aligned}
M_1 &= \sum_{x=0}^{C_1 - 1} x \frac{\mu^x}{x!} e^{-\mu} \\
&= \sum_{x=1}^{C_1 - 1} x \frac{\mu^x}{x!} e^{-\mu} \\
&= \mu \sum_{x=1}^{C_1 - 1} \frac{\mu^{x-1}}{(x-1)!} e^{-\mu} \\
&= \mu \sum_{y=0}^{C_1 - 2} \frac{\mu^y}{y!} e^{-\mu}
\end{aligned}
$$

and the last sum is just another Poisson probability, that is, M_1 can be calculated in R (with the obvious definitions of the variables) by

```
mu * ppois(c1 - 2, mu)
```

and M_2 by

```
mu * ppois(c2 - 1, mu, lower.tail = FALSE)
```

4.3 Negative Binomial

Let $X \sim \text{NegBin}(r, p)$ with $0 < p < 1$. Like R we consider the sample space to start at zero rather than r. This also allows for non-integer r. The densities of the family have the form

$$
f(x) = \frac{\Gamma(x + r)}{\Gamma(r) x!} p^r (1 - p)^x
$$

Note that if we are to have an exponential family r cannot be an unknown parameter! The only unknown parameter is p.

Again, all of the quantities in (15a) and (15b) are easily calculated except possibly M_1 and M_2. Do these work like the binomial and Poisson cases? Yes!

$$
\begin{aligned}
M_1 &= \sum_{x=0}^{C_1-1} x \frac{\Gamma(x+r)}{\Gamma(r) x!} p^r (1-p)^x \\
&= \sum_{x=1}^{C_1-1} x \frac{\Gamma(x+r)}{\Gamma(r) x!} p^r (1-p)^x \\
&= \sum_{y=0}^{C_1-2} \frac{\Gamma(y+1+r)}{\Gamma(r) y!} p^r (1-p)^{y+1} \\
&= \frac{1-p}{p} \sum_{y=0}^{C_1-2} \frac{\Gamma(y+1+r)}{\Gamma(r) y!} p^{r+1} (1-p)^y
\end{aligned}
$$

M_1 can be calculated in R (with the obvious definitions of the variables) by

```
(1 - p) / p * pnbinom(c2 - 2, r + 1, p)
```

and M_2 can be calculated by

```
(1 - p) / p * pnbinom(c1 - 1, r + 1, p, lower.tail = FALSE)
```

4.4 Two Independent Poisson Random Variables

Let $X_i \sim \text{Pois}(\mu_i)$ with $0 < \mu_i$, for $i = 1, 2$ be independent random variables. We wish to compare the means μ_1 and μ_2. We cannot just test or produce fuzzy confidence intervals for a function pulled out of the air, such as $\mu_1 - \mu_2$. The parameter we test must be canonical.

The canonical statistics of this exponential family are X_1 and X_2 and the corresponding canonical parameters are $\psi_i = \log(\mu_i)$. Linear functions of canonical parameters are again canonical so we can test or produce fuzzy confidence intervals for $\psi_1 - \psi_2 = \log(\mu_1/\mu_2)$.

Introduce new parameters

$$
\begin{aligned}
\psi_1 &= \eta + \theta \\
\psi_2 &= \eta
\end{aligned}
$$

Then

$$
X_1 \psi_1 + X_2 \psi_2 = X_1 \theta + (X_1 + X_2) \eta = T(\mathbf{X}) \theta + U(\mathbf{X}) \eta
$$

Where

$$
\begin{aligned}
T(\mathbf{X}) &= X_1 \\
U(\mathbf{X}) &= X_1 + X_2
\end{aligned} \tag{33}
$$

It is a standard result that the conditional distribution of $T(\mathbf{X})$ given $U(\mathbf{X})$ is

$$
X_1 \mid X_1 + X_2 \sim \text{Bin}\left(X_1 + X_2, \frac{\mu_1}{\mu_1 + \mu_2} \right)
$$

So the theory says we do the UMP or UMPU test based on this distribution with $\mu_1/(\mu_1+\mu_2)$ as the parameter of interest (Lehmann, TSH, 1st ed., pp. 140–142, gives further details).

4.5 Two Independent Binomial Random Variables

Let $X_i \sim \text{Bin}(n_i, p_i)$ with $0 < p_i < 1$, for $i = 1$, 2 be independent random variables. We wish to compare the proportions p_1 and p_2. We cannot just test or produce fuzzy confidence intervals for a function pulled out of the air, such as $p_1 - p_2$. The parameter we test must be canonical.

The canonical statistics of this exponential family are X_1 and X_2 and the corresponding canonical parameters are $\psi_i = \operatorname{logit}(p_i)$. Linear functions of canonical parameters are again canonical so we can test or produce fuzzy confidence intervals for $\psi_1 - \psi_2$.

As in the Poisson case we see that we can base the test on the conditional distribution of $T(\mathbf{X})$ given $U(\mathbf{X})$, where these variables are defined by (33). This distribution is (Lehmann, TSH, 1st ed., pp. 142–143) the exponential family generated by the hypergeometric distribution, which is called Fisher's noncentral hypergeometric distribution (Section 1.6 above). It has canonical parameter

$$\theta = \log\left(\frac{p_1(1-p_2)}{(1-p_1)p_2}\right)$$

So the theory says we do the UMP or UMPU test based on this distribution with θ as the parameter of interest

4.6 Two Independent Negative Binomial Variables

Let $X_i \sim \text{NegBin}(r_i, p_i)$ with $0 < r_i$ and $0 < p_i < 1$, for $i = 1$, 2 be independent random variables. As in Section 4.3 we are using the convention that the sample space starts at zero. We wish to compare the proportions p_1 and p_2. We cannot just test or produce fuzzy confidence intervals for a function pulled out of the air, such as $p_1 - p_2$. The parameter we test must be canonical.

The canonical statistics of this exponential family are X_1 and X_2 and the corresponding canonical parameters are $\psi_i = \log(1 - p_i)$. Linear functions of canonical parameters are again canonical so we can test or produce fuzzy confidence intervals for $\psi_1 - \psi_2$.

As in the Poisson case we see that we can base the test on the conditional distribution of $T(\mathbf{X})$ given $U(\mathbf{X})$, where these variables are defined by (33). This distribution is not fully explained in Lehmann, although the $r_1 = r_2 = 1$ case is the subject of a homework problem in the second edition.

Let's see what happens. The joint distribution of the X's is

$$
\begin{aligned}
f(x_1, x_2) &= \prod_{i=1}^{2} \frac{\Gamma(x_i + r_i)}{\Gamma(r_i) x_i!} p_i^{r_i} (1 - p_i)^{x_i} \\
&= \exp(x_1 \psi_1 + x_2 \psi_2) \prod_{i=1}^{2} \frac{\Gamma(x_i + r_i)}{\Gamma(r_i) x_i!} p_i^{r_i} \\
&= \exp(t\theta + u\eta) \frac{\Gamma(t + r_1)}{\Gamma(r_1) t!} \frac{\Gamma(u - t + r_2)}{\Gamma(r_2)(u - t)!} p_1^{r_1} p_2^{r_2} \qquad (34)
\end{aligned}
$$

where

$$
\begin{aligned}
t &= x_1 \\
u &= x_1 + x_2 \\
\theta &= \psi_1 - \psi_2 = \log\left(\frac{1 - p_1}{1 - p_2}\right) \\
\eta &= \psi_2 = \log(1 - p_2)
\end{aligned}
$$

We want to consider the conditional distribution of $T(\mathbf{X})$ given $U(\mathbf{X})$. Thought of as a function of t for fixed u and dropping all terms that do not contain t and θ we get

$$
f_\theta(t \mid u) = \frac{1}{c(\theta)} \exp(t\theta) \frac{\Gamma(t + r_1)\Gamma(u - t + r_2)}{t!(u - t)!}, \qquad t = 0, \ldots, u, \qquad (35)
$$

where $c(\theta)$ is chosen to make probabilities sum to one.

Equations (35) and (3) agree except the following changes of variable names

(35)	(3)
t	x
r_1	r
$u + r_2$	$N - r$
u	K

Thus we have indeed arrived at the exponential family described in Section 1.7 above.

4.7 Testing Independence in a Two-by-two Table

This is the UMP/UMPU competitor for Fisher's exact test. The data consist of a matrix X_{ij}, $i = 1, 2$, $j = 1, 2$, that has a multinomial distribution with sample size n and cell probability matrix p_{ij}, $i = 1, 2$, $j = 1, 2$. This is also called a two-by-two contingency table.

The canonical statistics are the X_{ij}, but the canonical parameters are not uniquely defined in terms of the p_{ij} because the model is really only three dimensional, not four, because the X_{ij} sum to n.

As is well known, this is a three-dimensional exponential family, the canonical statistics being any three of the four X_{ij}, the fourth being determined from the other three by the requirement that the X_{ij} sum to n.

In this problem (Lehmann, TSH, 1st ed., pp. 143–146) the marginals are the statistics for the nuisance parameters, and we can consider any other statistic linearly independent of the marginals and the sum of all cells as the statistic of interest. Lehmann chooses

$$\begin{aligned} T(\mathbf{X}) &= X_{11} \\ U_1(\mathbf{X}) &= X_{11} + X_{12} \\ U_2(\mathbf{X}) &= X_{11} + X_{21} \end{aligned} \tag{36}$$

The conditional distribution of $T(\mathbf{X})$ given the marginals $U_1(\mathbf{X})$ and $U_2(\mathbf{X})$ is well known. It is the hypergeometric distribution involved in Fisher's exact test under the null hypothesis of independence and under general null hypotheses is the exponential family generated by the hypergeometric we encountered in Sections 1.6 and 4.5. The canonical parameter is

$$\theta = \log\left(\frac{p_{11}p_{22}}{p_{12}p_{21}}\right)$$

4.8 Different Answers to the Same Question in a Poll

This section and the next give the UMP/UMPU/fuzzy competitors to the analysis of correlated binomial proportions in Wild and Seber (pp. 343–350). The first considers different answers to the same question on a poll. This is a multinomial problem. Say the categories of interest have counts X_1 and X_2, then we know

$$X_1 \mid X_1 + X_2 \sim \text{Bin}\left(X_1 + X_2, \frac{p_1}{p_1 + p_2}\right)$$

and so the UMP/UMPU/fuzzy procedures are based on this distribution.

4.9 Answers to Different Questions in a Poll

Here again, like in Section 4.7, we have a two-by-two table with data X_{ij} and cell probabilities p_{ij} but the question of interest is different. Now we are interested in just the opposite question, whether the marginals differ. This in a sense (a rather vague sense) interchanges the role of interest and nuisance parameters, what were interest parameters in Section 4.7 are now nuisance parameters and vice versa.

Ordinarily, this would be nonsense. There is exactly one interest parameter. The rest (in this case two) must be nuisance parameters. So, strictly speaking, they cannot be interchanged. But a two-by-two table has a redundant canonical statistic: there are four X_{ij} but they sum to n so only three are linearly independent. So if we add the redundant statistic to the statistics corresponding to parameters of interest we two sets of two that can be interchanged.

It is clear that (36) could have been written with subscripts 1 and 2 interchanged, which would make X_{21} the statistic of interest. This tells us that here we should condition on X_{11} and X_{22} leaving either X_{12} or X_{21} as the statistic of interest. Thus in this case the UMP/UMPU/fuzzy procedure is based on the distribution

$$X_{12} \mid X_{11}, X_{22} \sim \text{Bin}\left(n - X_{11} - X_{22}, \frac{p_{12}}{p_{12} + p_{21}}\right)$$

And in hindsight we see that we have invented the conditional, exact, UMP or UMPU competitor of McNemar's test (Lindgren, *Statistical Theory*, pp. 381–383).

5 Algorithms

5.1 Jobs

The obvious solution to our problems is to provide a function that calculates $\phi(x, \alpha, \theta)$ for each family of interest and for UMP and UMPU.

Probably such a function should follow the usual recycling rule, documented, for example, on the help page for `dbinom`, `pbinom`, `qbinom`, and `rbinom`

> The length of the result is determined by `n` for `rbinom`, and is the maximum of the lengths of the numerical arguments for the other functions.
>
> The numerical arguments other than `n` are recycled to the length of the result. Only the first elements of the logical arguments are used.

But rather than following this rule and only this rule, we should also return values at the knots (which users cannot specify because they don't know where they are).

- For fuzzy P-values, which are piecewise linear, there should at at least be an option to return only the knots and values at the knots, because that is all R needs to plot the function.

- Since probability density functions are more easily interpreted than cumulative distribution functions, there should also be an option to return the derivative of the fuzzy P-value, which is undefined at the knots and constant between knots. Of course, the derivative between two knots is given by rise over run (the difference of the values at those two knots divided by the difference of the knots, but that may involve catastrophic cancellation, and the computer should be more accurate).

- Fuzzy confidence intervals are constant on the interval where they are equal to one (the core) so we can save memory both for the returned R object, and any plots made from it, by not returning any values in the core (even if asked for by the user), at least as an option. Instead we return the endpoints of the core, and the values (one) at these points.

The set where a fuzzy confidence interval is nonzero is called its *support* (not to be confused with the concept of support in probability theory (where the probability mass function is nonzero). We can also save memory by returning the endpoints of the support, the values there (perhaps zero) and the boundaries of the parameter space and the values there (perhaps zero) and points and values for no other points outside the support.

Finally on the intervals where the fuzzy confidence intervals (where the membership function is nonlinear) we need to return the membership function values on a grid of parameter values and we need users to specify the density of points in the grid rather than the number of points because the users do not know the lengths of these intervals of rise and fall. Moreover, we need to insert the parameter values and membership function values at knots because users don't know those either, and not including the knots will make plots not show the knots.

When we return values of the function at points not specified by the user, we will also have to return those α values (for a fuzzy P-value) or θ values (for a fuzzy confidence interval). Thus we will return a list, and we might as well return a list specifying everything: $\phi(x, \alpha, \theta)$, x, α, θ, and the distribution and type of test (lower, upper, two-tailed), or, in the case of wanting the PDF of the fuzzy P-value, $\partial\phi(x, \alpha, \theta)/\partial\alpha$ rather than $\phi(x, \alpha, \theta)$.

We need make no special allowance for fuzzy decision functions $\phi(\,\cdot\,, \alpha, \theta)$, since the usual R behavior will accomodate that.

The package should also supply the usual d, p, q, and r functions (like `dbinom`, `pbinom`, `qbinom`, and `rbinom`) for the noncentral hypergeometric, and noncentral negative hypergeometric distributions.

5.2 Parameters

As seen in Algorithm 1 and various theorems, testing whether $x = \mu$ or not is important. We will not be able to do this with exact arithmetic if μ is calculated rather than provided by the user. Thus we should use μ as the parameter if at all possible.

- For the binomial distribution, we should use $\mu = np$ as the parameter rather than p, which R uses for `dbinom` and friends. If we have to multiply n times p, that will not be exact arithmetic, so the test may be in error.

- For the Poisson distribution, μ is already the usual parameter.

- For the negative binomial distribution, μ is already an optional parameter in `dnbinom` and friends. And the C interface provided to R packages has C functions `dnbinom_mu`, `pnbinom_mu`, `qnbinom_mu`, `rnbinom_mu` that use μ as a parameter.

- For the noncentral hypergeometric distribution we have the problem that we do not have a formula expressing the mapping between canonical and mean value parameters. So we may have to use the canonical parameter for it. Exception: for $\theta = 0$ we know $\mu = nK/N$ in the notation of Section 1.6, that is, when we have regular hypergeometric rather than noncentral hypergeometric. Since this is a widely used null hypothesis, we want to special case this.

- Similarly, for the noncentral negative hypergeometric distribution we also must use the canonical parameter, except when $\theta = 0$ when we know $\mu = rK/(N - K + 1)$ in the notation of Section 1.7.

REVISED DOWN TO HERE

5.3 Package Version 0.1

After a great struggle, a very simple algorithm was decided on for calculating $\phi(x, \alpha, \theta)$ for the binomial distribution. Given α and θ, calculate the appropriate c_1, c_2, γ_1, and γ_2 as follows.

First handle the special cases where α is zero or one and θ is on the boundary of the parameter space, using Theorem 6 above and the obvious fact that ϕ is identically equal to one when $\alpha = 1$ and identically equal to zero when $\alpha = 0$.

When in the general case $0 < \alpha < 1$ and θ not on the boundary choose some c_1 and c_2 such that $c_1 \leq E_\theta\{T(X)\} \leq c_2$. We pick the c_1 and c_2 that give, with randomization, an equal tailed test, in the hope that this is close to correct.

Then we go into an infinite loop that does the following.

- Calculate γ_1 and γ_2 using the current guesses for c_1 and c_2 and equations (15a) and (15b) above. If the results satisfy $0 \leq \gamma_1 \leq 1$ and $0 \leq \gamma_2 \leq 1$, then we are done and stop the loop.

- Otherwise, we change c_1 or c_2, the one corresponding to the γ_i that violates the constraints worst. If this γ_i is negative, we move the c_i out by one (i. e. decrease c_1 or increase c_2) and if this γ_i is greater than one, we move the c_i in by one.

Actually we don't do an infinite loop, because we have no theorem that says this algorithm converges, so we have a maximum iteration count (default 10) and just give up when it is reached. In the examples we have done, there has been no need to increase the iteration count.

See `ump/src/umpubinom.c` for an example of this algorithm. See `ump/tests/umpub.R` for the tests it passed.

5.4 Package Version 0.3

An attempt to implement the density of abstract randomized P-values and test the implementation shows that equations (15a) and (15b) are no good. They exhibit catastrophic cancellation for small alpha.

[block of text with several equations moved]

This seems to work better, but honesty compels us to admit that this formula also is subject to catastrophic cancellation. Perusal of the source code reveals several ad hoc bits of code that deal with special cases in which the code without the adhockery fails due to catastrophic cancellation or other problems with the inexactitude of floating point arithmetic.

It is fair to say that our code is far from an elegant and provably correct solution to this problem. We think Algorithm 1 would actually be better than the one we used in all respects except that it takes time proportional to the sample size, which was deemed unacceptable (perhaps wrongly).

5.5 Other Distributions

We could perhaps do Poisson and negative binomial similarly to the way version 0.3 does the binomial, but that will not do for the other two distributions of interest. For the noncentral hypergeometric we follow Liao and Rosen (2001) except they do not say how to calculate CDF, so we need to worry about that.

In fact, we think, all distributions of interest can be calculated following Liao and Rosen (2001) (whether we actually want to do that or not) so we discuss that.

5.5.1 Recursion

All of the distributions of interest satisfy a recursion relation. Define

$$r_\theta(x) = \frac{f_\theta(x)}{f_\theta(x-1)}$$

Then for the binomial exponential family of distributions

$$r_\theta(x) = \frac{\binom{n}{x} p^x (1-p)^{n-x}}{\binom{n}{x-1} p^{x-1} (1-p)^{n-x+1}} = \frac{(x-1)!\,(n-x+1)!\,p}{x!\,(n-x)!\,(1-p)} = \frac{(n-x+1)e^\theta}{x}$$

and for the Poisson exponential family of distributions

$$r_\theta(x) = \frac{\mu^x e^{-\mu}/x!}{\mu^{x-1} e^{-\mu}/(x-1)!} = \frac{\mu}{x} = \frac{e^\theta}{x}$$

and for the negative binomial exponential family of distributions

$$r_\theta(x) = \frac{\binom{r+x-1}{x} p^r (1-p)^x}{\binom{r+x-2}{x-1} p^r (1-p)^{x-1}} = \frac{(r+x-1)!\,(x-1)!\,(1-p)}{(r+x-2)!\,x!} = \frac{(r+x-1)e^\theta}{x}$$

and for the noncentral hypergeometric exponential family of distributions

$$
\begin{aligned}
r_\theta(x) &= \frac{e^{x\theta} \binom{K}{x} \binom{N-K}{n-x}}{e^{(x-1)\theta} \binom{K}{x-1} \binom{N-K}{n-x+1}} \\
&= \frac{(x-1)! \, (K-x+1)! \, (n-x+1)! (N-K-n+x-1)! \, e^\theta}{x! \, (K-x)! \, (n-x)! \, (N-K-n+x)!} \\
&= \frac{(K-x+1)(n-x+1) e^\theta}{x(N-K-n+x)}
\end{aligned}
$$

and for the noncentral negative hypergeometric exponential family of distributions

$$
\begin{aligned}
r_\theta(x) &= \frac{\binom{x+r-1}{x} \binom{N-r-x}{K-x} e^\theta}{\binom{x-1+r-1}{x-1} \binom{N-r-x+1}{K-x+1}} \\
&= \frac{(x+r-1)! \, (N-r-x)! \, (x-1)! \, (K-x+1)! \, e^\theta}{x! \, (K-x)! \, (x+r-2)! (N-r-x+1)!} \\
&= \frac{(x+r-1)(K-x+1) e^\theta}{x(N-r-x+1)}
\end{aligned}
$$

and these recursions are the fundamental idea of the method of Liao and Rosen (2001).

5.5.2 Unimodality

For binomial

$$
\frac{d}{dx} \log r_\theta(x) = -\frac{1}{n-x+1} - \frac{1}{x}
$$

(note that $n - x \geq 0$) so $r_\theta(x)$ is a decreasing function of x and all distributions in the family are unimodal.

For Poisson

$$
\frac{d}{dx} \log r_\theta(x) = -\frac{1}{x}
$$

so $r_\theta(x)$ is a decreasing function of x and all distributions in the family are unimodal.

For negative binomial

$$
\frac{d}{dx} \log r_\theta(x) = \frac{1}{r+x-1} - \frac{1}{x} = -\frac{r-1}{x(r+x-1)}
$$

(note that $r \geq 1$) so $r_\theta(x)$ is a decreasing function of x and all distributions in the family are unimodal.

For noncentral hypergeometric

$$
\frac{d}{dx} \log r_\theta(x) = -\frac{1}{K-x+1} - \frac{1}{n-x+1} - \frac{1}{x} - \frac{1}{N-K-n+x}
$$

(note the denominator of each fraction is positive) so $r_\theta(x)$ is a decreasing function of x and all distributions in the family are unimodal.

For noncentral negative hypergeometric

$$
\begin{aligned}
\frac{d}{dx} \log r_\theta(x) &= \frac{1}{x+r-1} - \frac{1}{K-x+1} - \frac{1}{x} + \frac{1}{N-r-x+1} \\
&= -\frac{r-1}{x(x+r-1)} - \frac{N-K-r}{(K-x+1)(N-r-x+1)}
\end{aligned}
$$

(note that $1 \le r \le N-K$ and $1 \le x \le K$) so $r_\theta(x)$ is a decreasing function of x and all distributions in the family are unimodal.

5.5.3 Modes

Thus for each distribution the largest x such that $r_\theta(x) \ge 1$ is the unique mode if $r_\theta(x) > 1$ for that x, whereas x and $x-1$ are both modes if $r_\theta(x) = 1$ for that x.

To simplify notation, for all distributions, we write $\rho = e^\theta$ and we write m_θ for the mode.

For binomial

$$
\begin{aligned}
r_\theta(x) &= \frac{e^\theta}{x} \\
m_\theta &= \lfloor \rho \rfloor
\end{aligned}
$$

For Poisson

$$
\begin{aligned}
r_\theta(x) &= \frac{(n-x+1)e^\theta}{x} \\
m_\theta &= \lfloor (n+1)\rho/(1+\rho) \rfloor
\end{aligned}
$$

For negative binomial

$$
\begin{aligned}
r_\theta(x) &= \frac{(r+x-1)e^\theta}{x} \\
m_\theta &= \lfloor (r-1)\rho/(1-\rho) \rfloor
\end{aligned}
$$

For noncentral hypergeometric

$$
r_\theta(x) = \frac{(K-x+1)(n-x+1)e^\theta}{x(N-K-n+x)}
$$

so to find the mode we need to solve the quadratic equation

$$
(K-x+1)(n-x+1)\rho = x(N-K-n+x)
$$

for x. This equation can be rewritten

$$
\rho\bigl(x^2 - (K+n+2)x + (K+1)(n+1)\bigr) = x^2 + (N-K-n)x
$$

or

$$
(\rho-1)x^2 - \bigl((K+n)(\rho-1) + N + 2\rho\bigr)x + \rho(K+1)(n+1) = 0
$$

References

Agresti, A. (1992). A survey of exact inference for contingency tables. *Statistical Science*, **7**, 131–153. doi:10.1214/ss/1177011454.

Barndorff-Nielsen, O. (1978). *Information and Exponential Families.* Wiley, Chichester.

Cornfield, J. (1956). A statistical problem arising from retrospective studies. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, **4**, 135–148. University of California, Berkeley.

Ferguson, T. S. (1996). *A Course in Large Sample Theory.* London: Chapman & Hall.

Fisher, R. A. (1935). The logic of inductive inference (with discussion). *Journal of the Royal Statistical Society, Series A*, **98**, 39–82. doi:10.2307/2342435.

Geyer, C. J. (1990). *Likelihood and Exponential Families.* PhD thesis, University of Washington. `http://purl.umn.edu/56330`.

Geyer, C. J. and Meeden, G. D. (2005). Fuzzy and randomized confidence intervals and *P*-values (with discussion). *Statistical Science*, **20**, 358–387. doi:10.1214/088342305000000340.

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

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

Liao, J. G., and Rosen, O. (2001). Fast and stable algorithms for computing and sampling from the noncentral hypergeometric distribution. *American Statistician*, **55**, 366–369. doi:10.1198/000313001753272547.

Martin, A. D., Quinn, K. M., Park, J. H. (2022). R package `MCMCpack`: : Markov chain Monte Carlo (MCMC) package, version 1.6-3. `https://cran.r-project.org/package=MCMCpack`.

Wikipedia contributors (2021). Fisher's noncentral hypergeometric distribution. In *Wikipedia, The Free Encyclopedia*. `https://en.wikipedia.org/w/index.php?title=Fisher%27s_noncentral_hypergeometric_distribution&oldid=1004931849`.

Wikipedia contributors (2023a). Hypergeometric distribution. In *Wikipedia, The Free Encyclopedia*. `https://en.wikipedia.org/w/index.php?title=Hypergeometric_distribution&oldid=1162227027`.

Wikipedia contributors (2023b). Negative hypergeometric distribution. In *Wikipedia, The Free Encyclopedia*. `https://en.wikipedia.org/w/index.php?title=Negative_hypergeometric_distribution&oldid=1160615854`.

A Algorithm for Fuzzy P-Values

Here we are interested in x fixed at the observed data and θ fixed at the value hypothesized under the null hypothesis.

We know from Algorithm 1 that $C_1 \le \mu \le C_2$. Hence if $T(x) \le \mu$, we have $\phi(x, \alpha, \theta)$ strictly between 0 and 1 if and only if $C_1 = T(x)$.

Then in (14a) and (14b) p_1 and P_1 and M_1 are fixed (because C_1 and θ are fixed), from which we see

$$
\begin{aligned}
P_1 + \gamma_1 p_1 &\le \alpha \\
M_1 + \gamma_1 C_1 p_1 &\le \alpha\mu
\end{aligned}
$$

so

$$
\alpha \ge \max\left(P_1, \frac{M_1}{\mu}\right)
$$

So now we need a guess about what C_2 may be. Because of discreteness, we have no way to calculate except by trial and error. Start with an approximately equal-tailed interval, that is,

$$
\begin{aligned}
P_1 &\approx P_2 \\
P_1 + p_1 &\approx p_2 + P_2
\end{aligned}
$$

For example, we could set C_2 to be the $1 - P_1$ quantile.

The next step is to use (15a) and (15b) to see if we have a solution, that is, if there is any range of α for which these equations both evaluate to something between 0 and 1, that is, α such that

$$
\begin{aligned}
0 &\le \frac{\alpha(C_2 - \mu) + (M_1 - C_2 P_1) + (M_2 - C_2 P_2)}{p_1(C_2 - C_1)} \le 1 \\
0 &\le \frac{\alpha(\mu - C_1) - (M_2 - C_1 P_2) - (M_1 - C_1 P_1)}{p_2(C_2 - C_1)} \le 1
\end{aligned}
$$

both hold. We should put in the assumption that $C_1 < C_2$. In case we have $T(x) = \mu$, the fuzzy P-value is given by (12a) with $C_1 = C_2 = C = \mu$ and $p_1 = p_2 = p$.

$$
\begin{aligned}
\frac{-(M_1 - C_2 P_1) - (M_2 - C_2 P_2)}{C_2 - \mu} &\le \alpha \le \frac{p_1(C_2 - C_1) - (M_1 - C_2 P_1) - (M_2 - C_2 P_2)}{C_2 - \mu} \\
\frac{(M_2 - C_1 P_2) + (M_1 - C_1 P_1)}{\mu - C_1} &\le \alpha \le \frac{p_2(C_2 - C_1) + (M_2 - C_1 P_2) + (M_1 - C_1 P_1)}{\mu - C_1}
\end{aligned}
$$

Try an example, Poisson, execute algorithm 1.

```
fuzzy.pval <- function(mu) {
    stopifnot(is.numeric(mu))
```

```r
    stopifnot(is.finite(mu))
    stopifnot(length(mu) == 1)
    stopifnot(mu > 0)
    C1 <- floor(mu)
    C2 <- ceiling(mu)
    save <- c(C1, C2, 1, 1, 1) |> rbind(deparse.level = 0)
    colnames(save) <- c("C1", "C2", "gamma1", "gamma2", "alpha")
    if (C1 == C2) {
        p <- ppois(C1, mu)
        alpha <- 1 - p
        gamma <- 1 - (1 - alpha) / p
        save <- rbind(save, c(C1, C2, gamma, gamma, alpha))
        C1 <- C1 - 1
        C2 <- C2 + 1
        save <- rbind(save, c(C1, C2, 1, 1, alpha))
    }
    gamma1 <- 1
    gamma2 <- 1
    repeat {
        p1 <- dpois(C1, mu)
        p2 <- dpois(C2, mu)
        P1 <- ppois(C1 - 1, mu)
        P2 <- ppois(C2, mu, lower.tail = FALSE)
        M1 <- mu * ppois(C1 - 2, mu)
        M2 <- mu * ppois(C2 - 1, mu, lower.tail = FALSE)
        L1 <- (- (M1 - C2 * P1) - (M2 - C2 * P2)) / (C2 - mu)
        L2 <- ((M2 - C1 * P2) + (M1 - C1 * P1)) / (mu - C1)
        alpha <- max(L1, L2)
        gamma1 <- (alpha * (C2 - mu) + (M1 - C2 * P1) + (M2 - C2 * P2)) /
            (p1 * (C2 - C1))
        gamma2 <- (alpha * (mu - C1) - (M2 - C1 * P2) - (M1 - C1 * P1)) /
            (p2 * (C2 - C1))
        save <- rbind(save, c(C1, C2, gamma1, gamma2, alpha))
        if (L1 > L2) {
            C1 <- C1 - 1
            gamma1 <- 1
        } else {
            C2 <- C2 + 1
            gamma2 <- 1
        }
        save <- rbind(save, c(C1, C2, gamma1, gamma2, alpha))
        if (alpha < 1e-5) break
    }
    as.data.frame(save)
}
```

```r
fuzzy.pval(5.55) |> zapsmall()
```

```
##    C1 C2   gamma1   gamma2    alpha
## 1   5  6 1.000000 1.000000 1.000000
## 2   5  6 0.243182 0.000000 0.713101
## 3   5  7 0.243182 1.000000 0.713101
## 4   5  7 0.000000 0.874227 0.655882
## 5   4  7 1.000000 0.874227 0.655882
## 6   4  7 0.334237 0.000000 0.444193
## 7   4  8 0.334237 1.000000 0.444193
## 8   4  8 0.000000 0.625582 0.360329
## 9   3  8 1.000000 0.625582 0.360329
## 10  3  8 0.529017 0.000000 0.253865
## 11  3  9 0.529017 1.000000 0.253865
## 12  3  9 0.000000 0.190822 0.151960
## 13  2  9 1.000000 0.190822 0.151960
## 14  2  9 0.834218 0.000000 0.131821
## 15  2 10 0.834218 1.000000 0.131821
## 16  2 10 0.212287 0.000000 0.064880
## 17  2 11 0.212287 1.000000 0.064880
## 18  2 11 0.000000 0.447612 0.043891
## 19  1 11 1.000000 0.447612 0.043891
## 20  1 11 0.627556 0.000000 0.029147
## 21  1 12 0.627556 1.000000 0.029147
## 22  1 12 0.172113 0.000000 0.012388
## 23  1 13 0.172113 1.000000 0.012388
## 24  1 13 0.000000 0.233640 0.006407
## 25  0 13 1.000000 0.233640 0.006407
## 26  0 13 0.761252 0.000000 0.004788
## 27  0 14 0.761252 1.000000 0.004788
## 28  0 14 0.301782 0.000000 0.001828
## 29  0 15 0.301782 1.000000 0.001828
## 30  0 15 0.111659 0.000000 0.000655
## 31  0 16 0.111659 1.000000 0.000655
## 32  0 16 0.038732 0.000000 0.000221
## 33  0 17 0.038732 1.000000 0.000221
## 34  0 17 0.012645 0.000000 0.000070
## 35  0 18 0.012645 1.000000 0.000070
## 36  0 18 0.003899 0.000000 0.000021
## 37  0 19 0.003899 1.000000 0.000021
## 38  0 19 0.001139 0.000000 0.000006
## 39  0 20 0.001139 1.000000 0.000006
```

```r
fuzzy.pval(5) |> zapsmall()
```

```
##    C1 C2   gamma1   gamma2    alpha
```

```
## 1   5  5 1.000000 1.000000 1.000000
## 2   5  5 0.000000 0.000000 0.384039
## 3   4  6 1.000000 1.000000 0.384039
## 4   4  6 0.166667 0.000000 0.532087
## 5   4  7 0.166667 1.000000 0.532087
## 6   4  7 0.000000 0.860000 0.488220
## 7   3  7 1.000000 0.860000 0.488220
## 8   3  7 0.360119 0.000000 0.308575
## 9   3  8 0.360119 1.000000 0.308575
## 10  3  8 0.000000 0.483733 0.224323
## 11  2  8 1.000000 0.483733 0.224323
## 12  2  8 0.625083 0.000000 0.161168
## 13  2  9 0.625083 1.000000 0.161168
## 14  2  9 0.050972 0.000000 0.076549
## 15  2 10 0.050972 1.000000 0.076549
## 16  2 10 0.000000 0.857946 0.069680
## 17  1 10 1.000000 0.857946 0.069680
## 18  1 10 0.422786 0.000000 0.034677
## 19  1 11 0.422786 1.000000 0.034677
## 20  1 11 0.055812 0.000000 0.014071
## 21  1 12 0.055812 1.000000 0.014071
## 22  1 12 0.000000 0.687136 0.011117
## 23  0 12 1.000000 0.687136 0.011117
## 24  0 12 0.509686 0.000000 0.005453
## 25  0 13 0.509686 1.000000 0.005453
## 26  0 13 0.196033 0.000000 0.002019
## 27  0 14 0.196033 1.000000 0.002019
## 28  0 14 0.070012 0.000000 0.000698
## 29  0 15 0.070012 1.000000 0.000698
## 30  0 15 0.023337 0.000000 0.000226
## 31  0 16 0.023337 1.000000 0.000226
## 32  0 16 0.007293 0.000000 0.000069
## 33  0 17 0.007293 1.000000 0.000069
## 34  0 17 0.002145 0.000000 0.000020
## 35  0 18 0.002145 1.000000 0.000020
## 36  0 18 0.000596 0.000000 0.000005
## 37  0 19 0.000596 1.000000 0.000005
```

Looks like that works. Now we need a function that works for just one x.

```
fuzzy.pval.too <- function(mu, x) {
    stopifnot(is.numeric(mu))
    stopifnot(is.finite(mu))
    stopifnot(length(mu) == 1)
    stopifnot(mu > 0)
    stopifnot(is.numeric(x))
```

```
    stopifnot(is.finite(x))
    stopifnot(length(x) == 1)
    stopifnot(x >= 0)
    stopifnot(x == round(x))

    # special case x == mu
    if (x == mu) {
        p <- ppois(x, mu)
        alpha <- 1 - p
        return(list(alpha = c(alpha, 1), df = c(0, 1)))
    }

    # regular case mu < x
    if (x > mu) {
        C2 <- x
        p2 <- dpois(C2, mu)
        P2 <- ppois(C2, mu, lower.tail = FALSE)
        M2 <- mu * ppois(C2 - 1, mu, lower.tail = FALSE)
        # guess C1
        C1 <- qpois(P2, mu)
        p1 <- dpois(C1, mu)
        P1 <- ppois(C1 - 1, mu)
        M1 <- mu * ppois(C1 - 2, mu)
        L1 <- (- (M1 - C2 * P1) - (M2 - C2 * P2)) / (C2 - mu)
        L2 <- ((M2 - C1 * P2) + (M1 - C1 * P1)) / (mu - C1)
        U1 <- (p1 * (C2 - C1) - (M1 - C2 * P1) - (M2 - C2 * P2)) / (C2 - mu)
        U2 <- (p2 * (C2 - C1) + (M2 - C1 * P2) + (M1 - C1 * P1)) / (mu - C1)
        alpha <- c(max(L1, L2), min(U1, U2))
        gamma2 <- (alpha * (mu - C1) - (M2 - C1 * P2) - (M1 - C1 * P1)) /
            (p2 * (C2 - C1))
        gamma2 <- zapsmall(gamma2)
        return(list(alpha = alpha, df = gamma2))
    }
}

fuzzy.pval.too(5, 5)

## $alpha
## [1] 0.3840393 1.0000000
##
## $df
## [1] 0 1

fuzzy.pval(5.55) |> subset(C2 == 8) |> zapsmall()

##    C1 C2    gamma1    gamma2     alpha
```

```
## 7    4   8 0.334237 1.000000 0.444193
## 8    4   8 0.000000 0.625582 0.360329
## 9    3   8 1.000000 0.625582 0.360329
## 10   3   8 0.529017 0.000000 0.253865
```

```
fuzzy.pval.too(5.55, 8)
```

```
## $alpha
## [1] 0.2538648 0.3603289
##
## $df
## [1] 0.0000000 0.6255818
```

B Fuzzy Confidence Interval Examples

First example (Figure 1 below), note that when observed x is as low as possible (zero for binomial), the fuzzy confidence interval starts at $1 - \alpha$ as per Theorem 6. Also note that in this case the core (set where membership function is equal to one) is empty.

```
library(ump)
p <- seq(0, 1, length = 1001)
phi <- umpu.binom(0, 10, p, 0.05)
xmax <- min(p[phi == 1])
xmax <- 1.1 * xmax
plot(p, 1 - phi, xlim = c(0, xmax), ylim = 0:1,
    xlab = "success probability", ylab = "", type = "l")
```

Second example (Figure 2 below), note that when observed x is adjacent to its minimum value (one for binomial), the fuzzy confidence interval starts at $1 - \alpha$ as per Theorem 6 and goes up from there until it reaches one, and is nonincreasing thereafter. In this example, the core is nonempty.

```
phi <- umpu.binom(1, 10, p, 0.05)
xmax <- min(p[phi == 1])
xmax <- 1.1 * xmax
plot(p, 1 - phi, xlim = c(0, xmax), ylim = 0:1,
    xlab = "success probability", ylab = "", type = "l")
```

Third example (Figure 3 below), another one just like the previous one except that the coverage probability is less, so the behavior is more extreme.

```
phi <- umpu.binom(1, 10, p, 0.5)
xmax <- min(p[phi == 1])
```

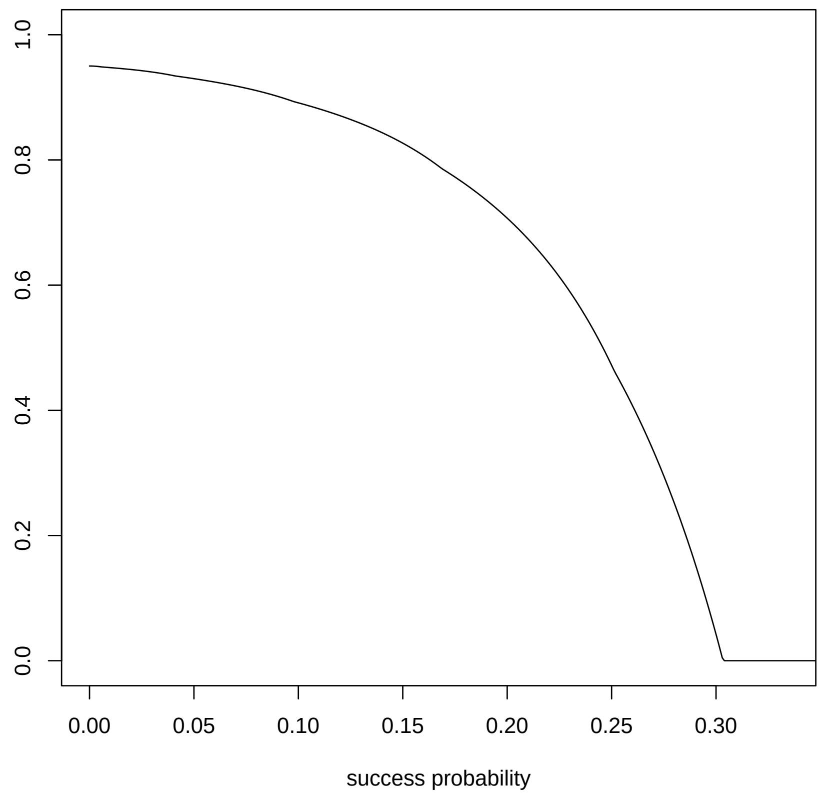

Figure 1: Binomial Fuzzy Confidence Interval, x = 0, n = 10, alpha = 0.05

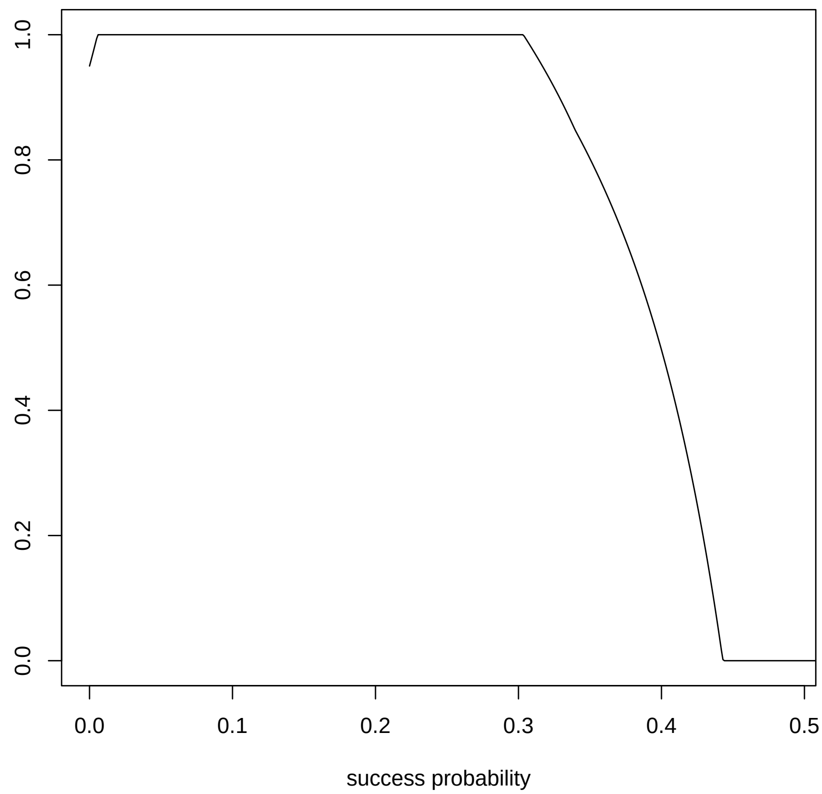

Figure 2: Binomial Fuzzy Confidence Interval, x = 1, n = 10, alpha = 0.05

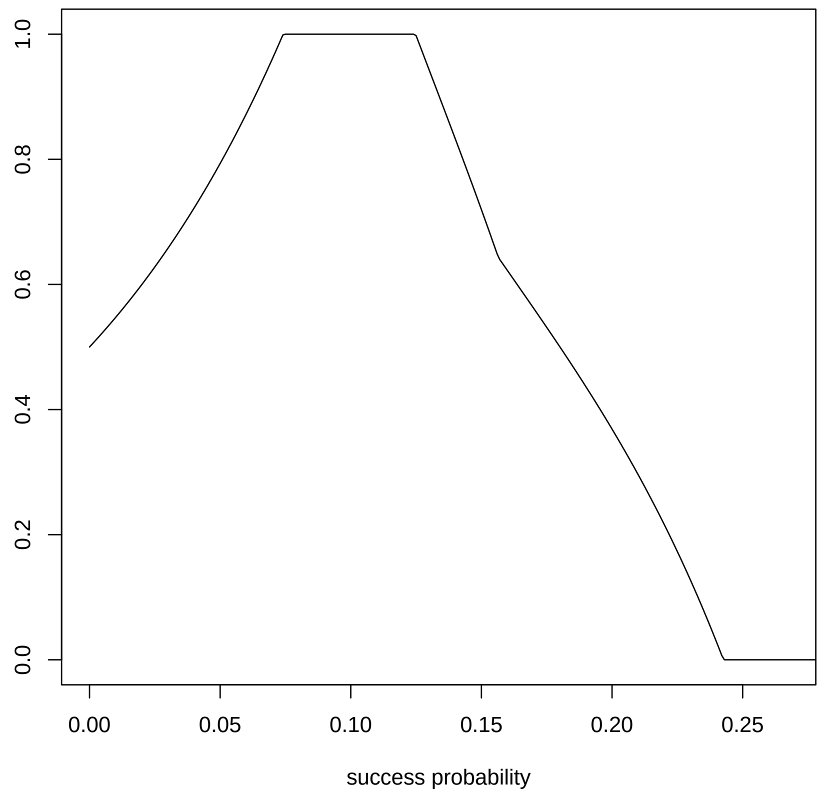

Figure 3: Binomial Fuzzy Confidence Interval, x = 1, n = 10, alpha = 0.5

```
xmax <- 1.1 * xmax
plot(p, 1 - phi, xlim = c(0, xmax), ylim = 0:1,
    xlab = "success probability", ylab = "", type = "l")
```

Fourth example (Figure 4 below), another one just like the previous one except that the coverage probability is less again, so the behavior is even more extreme. For this one the coverage probability requested is so low that the core is empty. In order to have exactly 25% confidence, the membership function cannot go up to one. The membership function reaches its maximum value at $\mu = x$ as per Theorem 5.

```
phi <- umpu.binom(1, 10, p, 0.75)
xmax <- min(p[phi == 1])
```

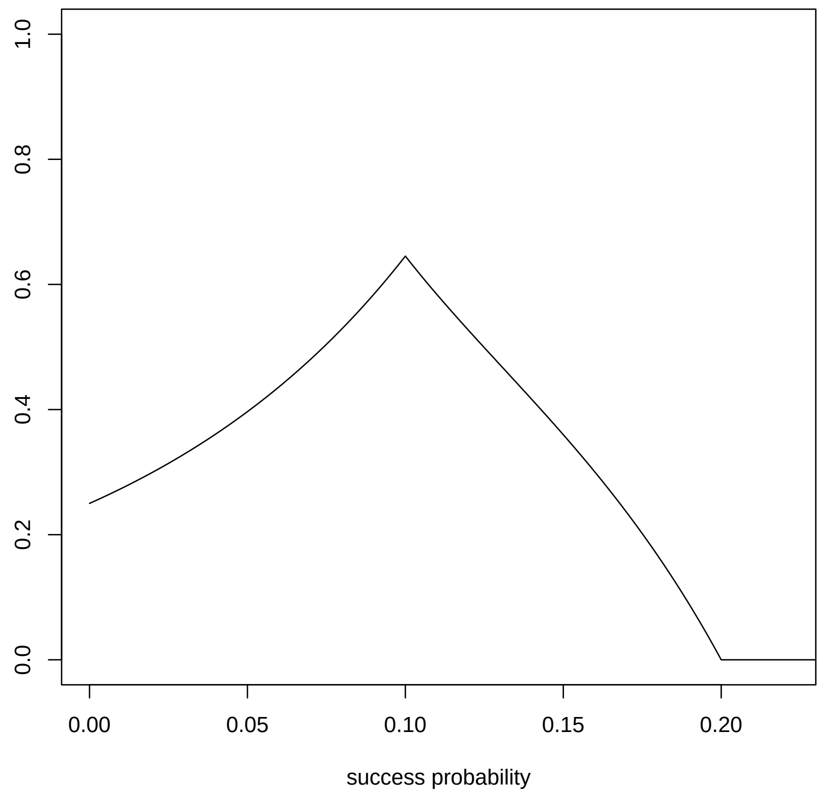

Figure 4: Binomial Fuzzy Confidence Interval, x = 1, n = 10, alpha = 0.75

```
xmax <- 1.1 * xmax
plot(p, 1 - phi, xlim = c(0, xmax), ylim = 0:1,
    xlab = "success probability", ylab = "", type = "l")
```

Fifth example (Figure 5 below), now the observed x is not either of the two lowest possible values or either of the two highest. So the fuzzy confidence interval has the typical behavior: as μ and θ increase from one end of the parameter space to the other, the fuzzy confidence interval is zero for a while, then increases smoothly to one, then is one for a while (the core), then decreases smoothly to zero, then is zero the rest of the way.

```
phi <- umpu.binom(4, 10, p, 0.05)
xmax <- max(p[phi < 1])
```

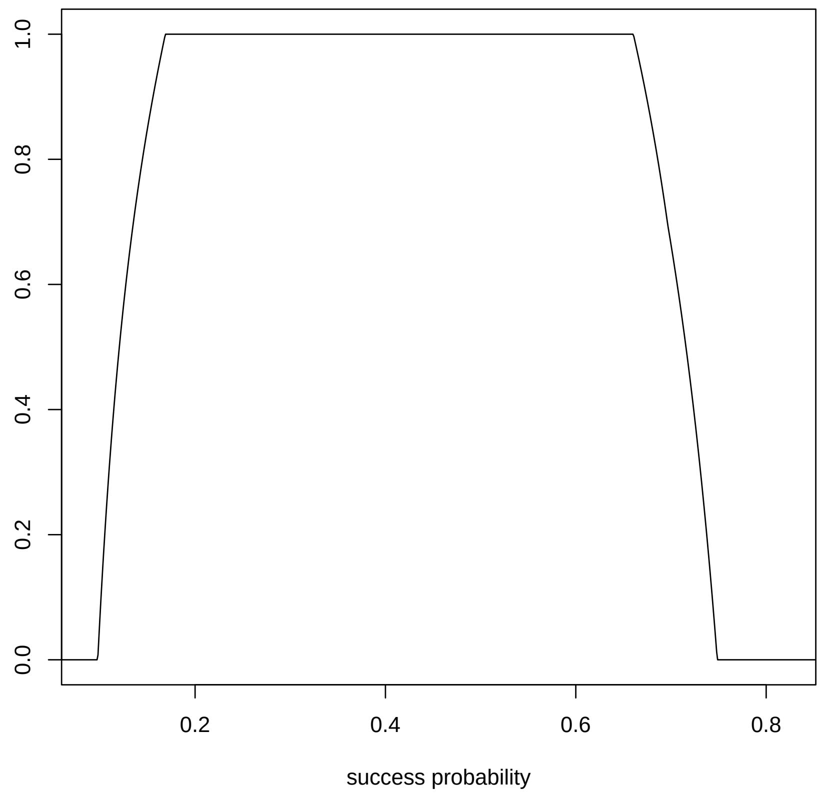

Figure 5: Binomial Fuzzy Confidence Interval, x = 4, n = 10, alpha = 0.05

```
xmin <- min(p[phi < 1])
xmax <- 1.1 * xmax
xmin <- xmin / 1.1
plot(p, 1 - phi, xlim = c(xmin, xmax), ylim = 0:1,
    xlab = "success probability", ylab = "", type = "l")
```

Sixth example (Figure 6 below), another one just like the previous one except that the coverage probability is less, so the behavior is more extreme.

```
phi <- umpu.binom(4, 10, p, 0.5)
xmax <- max(p[phi < 1])
xmin <- min(p[phi < 1])
xmax <- 1.1 * xmax
```

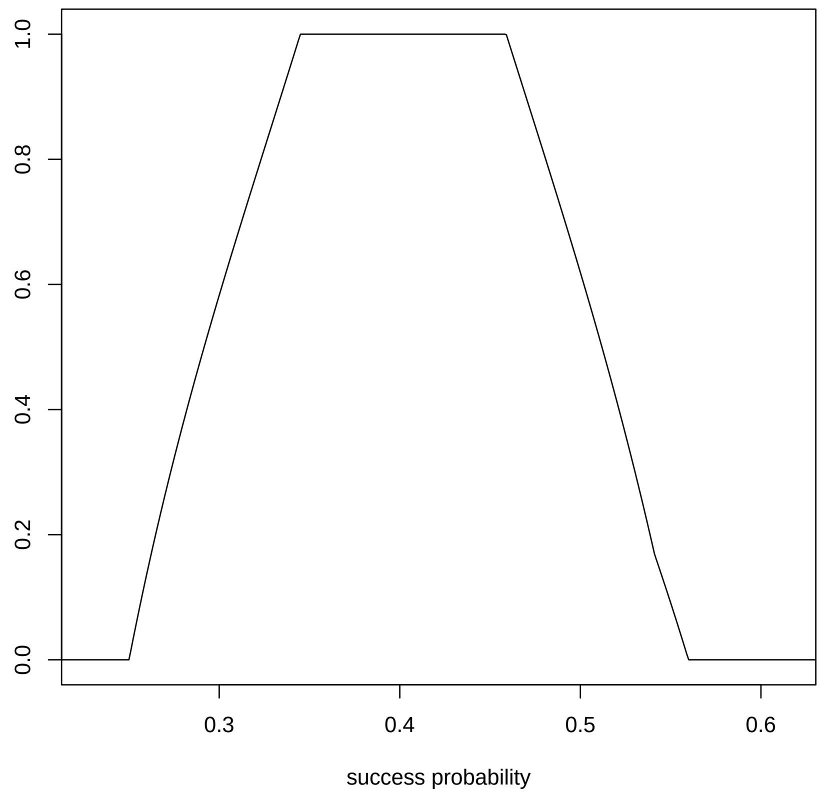

Figure 6: Binomial Fuzzy Confidence Interval, x = 4, n = 10, alpha = 0.5

```
xmin <- xmin / 1.1
plot(p, 1 - phi, xlim = c(xmin, xmax), ylim = 0:1,
    xlab = "success probability", ylab = "", type = "l")
```

Seventh example (Figure 7 below), another one just like the previous one except that the coverage probability is less again, so the behavior is even more extreme. For this one the coverage probability requested is so low that the core is empty. In order to have exactly 20% confidence, the membership function cannot go up to one. The membership function reaches its maximum value at $\mu = x$ as per Theorem 5.

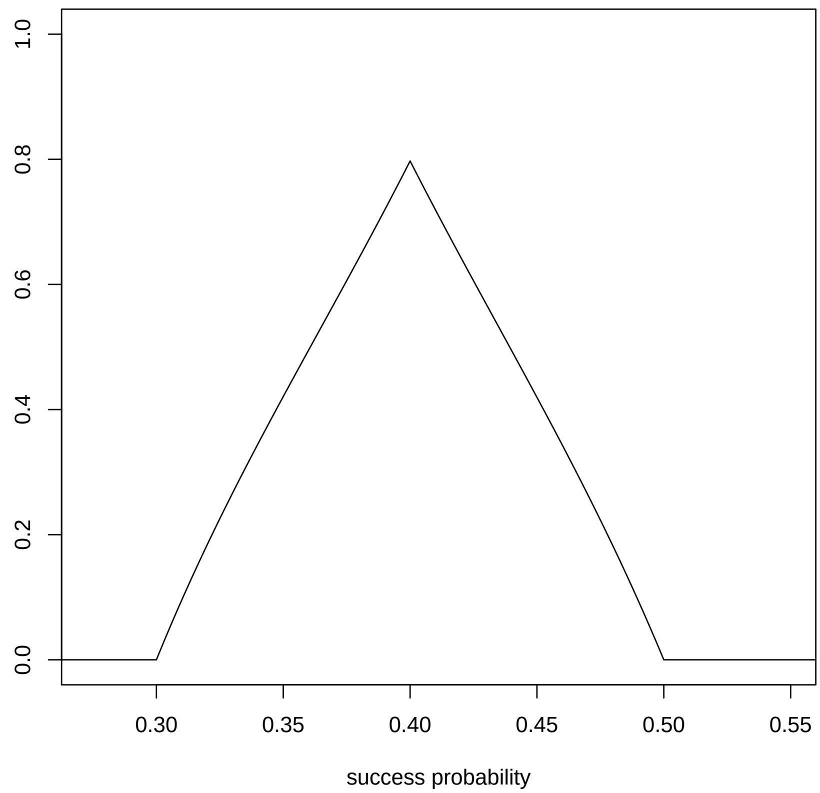

Figure 7: Binomial Fuzzy Confidence Interval, x = 4, n = 10, alpha = 0.8

```
phi <- umpu.binom(4, 10, p, 0.8)
xmax <- max(p[phi < 1])
xmin <- min(p[phi < 1])
xmax <- 1.1 * xmax
xmin <- xmin / 1.1
plot(p, 1 - phi, xlim = c(xmin, xmax), ylim = 0:1,
    xlab = "success probability", ylab = "", type = "l")
```