Fuzzy Confidence Intervals and P-values

Charles Geyer
University of Minnesota

joint work with

Glen Meeden
University of Minnesota

`http://www.stat.umn.edu/geyer/fuzz`

Ordinary Confidence Intervals

OK for continuous data, but a really bad idea for discrete data.

Why?

Coverage Probability

$$\begin{aligned}\gamma(\theta) &= \mathrm{pr}_\theta\{l(X) < \theta < u(X)\} \\ &= \sum_{x \in S} I_{(l(x), u(x))}(\theta) \cdot f_\theta(x)\end{aligned}$$

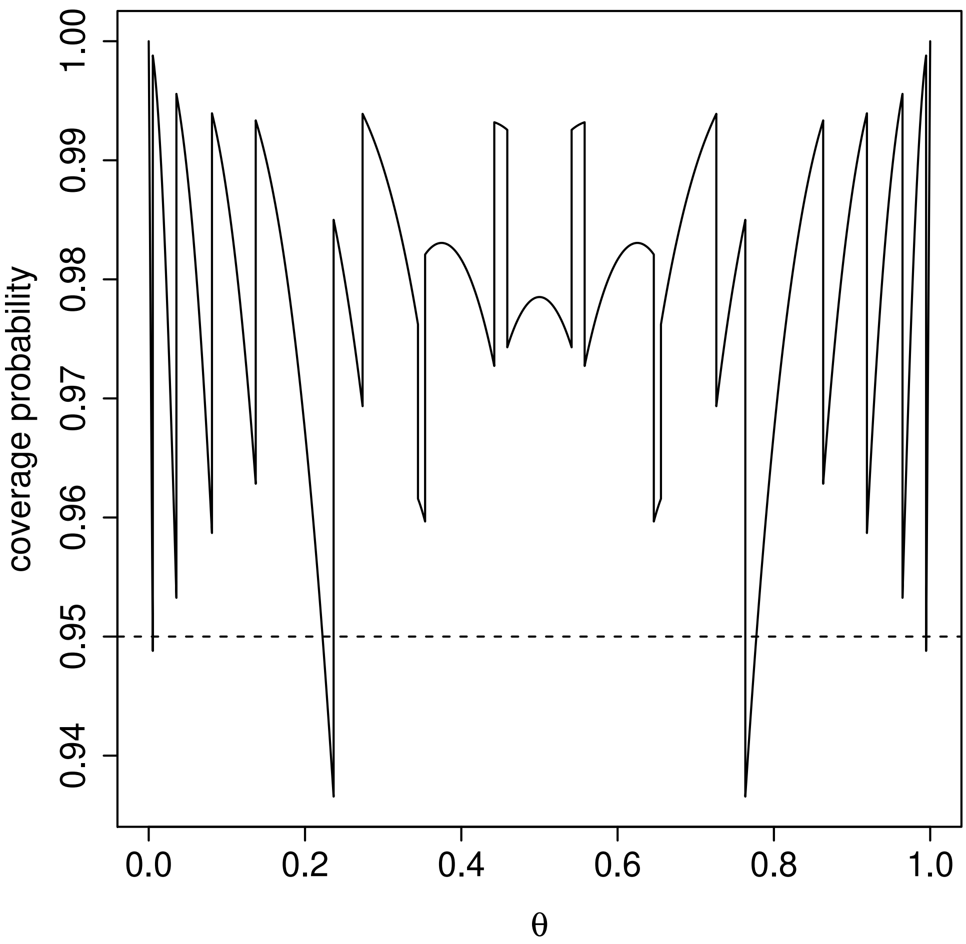
As θ moves across the boundary of a possible confidence interval $(l(x), u(x))$, the coverage probability jumps by $f_\theta(x)$.

Ideally, γ is a constant function equal to the nominal confidence coefficient.

But that's not possible.

Binomial Example

Binomial data, sample size $n = 10$, confidence interval calculated by R function `prop.test`



Recent Literature

Agresti and Coul (*Amer. Statist.*, 1998)
Approximate is better than "exact" for interval estimation of binomial proportions.

Brown, Cai, and DasGupta (*Statist. Sci.*, 2001)
Interval estimation for a binomial proportion (with discussion).

Casella (*Statist. Sci.*, 2001)
Comment on Brown, et al.

All recommend different intervals. All recommended intervals are bad, just slightly less bad than other possibilities.

Ordinary confidence intervals for discrete data are irreparably bad.

Randomized Tests

Randomized test defined by *critical function* $\phi(x, \alpha, \theta)$.

- observed data x.
- significance level α
- null hypothesis $H_0 : \theta = \theta_0$

Decision is randomized: reject H_0 with probability $\phi(x, \alpha, \theta_0)$.

Since probabilities are between zero and one, so is $\phi(x, \alpha, \theta)$.

Classical uniformly most powerful (UMP) and UMP unbiased (UMPU) tests are randomized when data are discrete.

Randomized Test Example

Observe $X \sim \text{Binomial}(20, \theta)$. Test

$$H_0 : \theta = 0.5$$
$$H_1 : \theta < 0.5$$

Distribution of X under H_0.

x	f(x)	F(x)
4	0.0046	0.0059
5	0.0148	0.0207
6	0.0370	0.0577
7	0.0739	0.1316

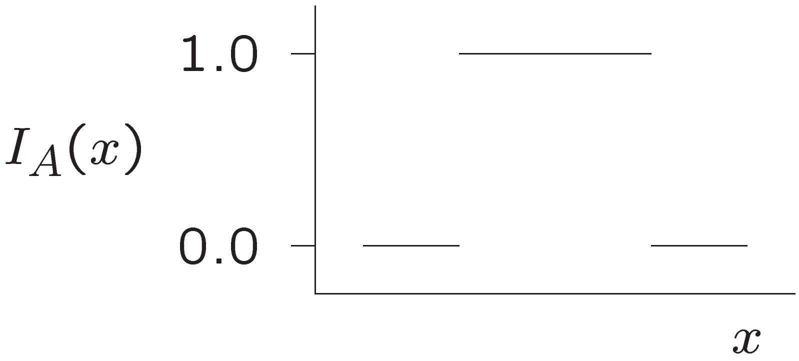
Nonrandomized test can have $\alpha = 0.0207$ or $\alpha = 0.0577$, but nothing in between.

Randomized test that rejects with probability one when $X \leq 5$ and with probability 0.7928 when $X = 6$ has $\alpha = 0.05$.

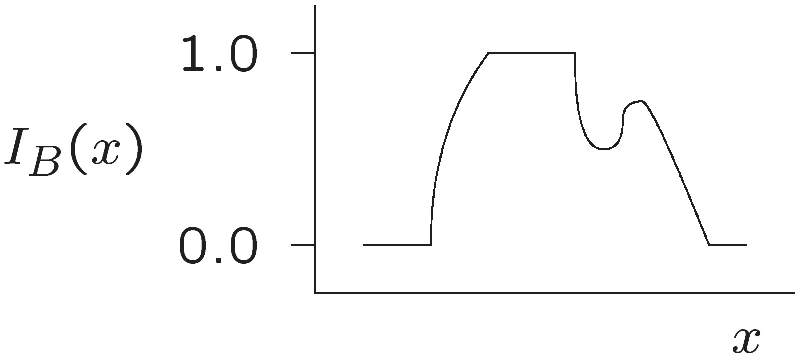
$$\begin{aligned} &\Pr(X \leq 5) + 0.7928 \cdot \Pr(X = 6) \\ &= 0.0207 + 0.7928 \cdot 0.0370 \\ &= 0.0500 \end{aligned}$$

Fuzzy Sets

Indicator function I_A of ordinary set A



Membership function I_B of fuzzy set B



Membership function $I_B(x)$ indicates degree to which x is to be considered to be in set B.

Ordinary sets are special case of fuzzy sets called *crisp sets*.

Fuzzy Tests and Confidence Intervals

- For fixed α and θ_0,

  $$x \mapsto \phi(x, \alpha, \theta_0)$$

  is the *fuzzy decision function* for the size α test of $H_0 : \theta = \theta_0$.

- For fixed x and α,

  $$\theta \mapsto 1 - \phi(x, \alpha, \theta)$$

  is (the membership function of) the *fuzzy confidence interval* with coverage $1 - \alpha$.

- For fixed x and θ_0,

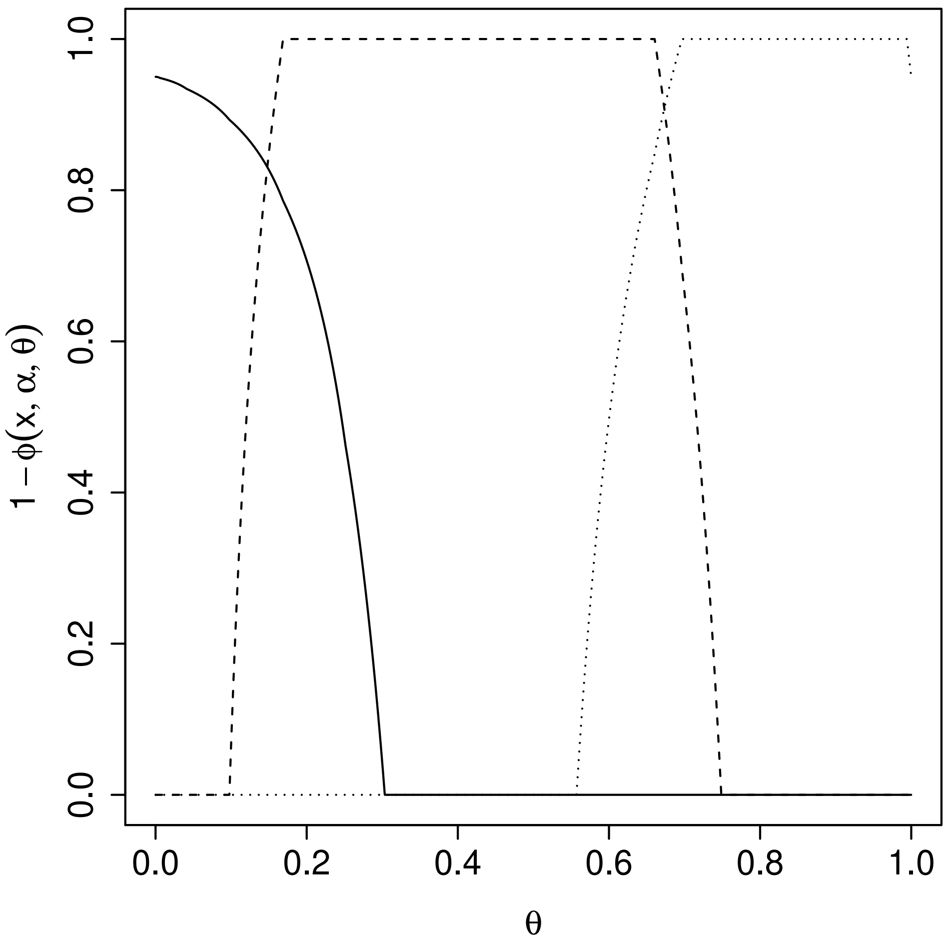
  $$\alpha \mapsto \phi(x, \alpha, \theta_0)$$

  is (the cumulative distribution function of) the *fuzzy P-value* for test of $H_0 : \theta = \theta_0$.

Binomial Example

Sample size $n = 10$, fuzzy confidence interval associated with UMPU test, confidence level $1 - \alpha = 0.95$.

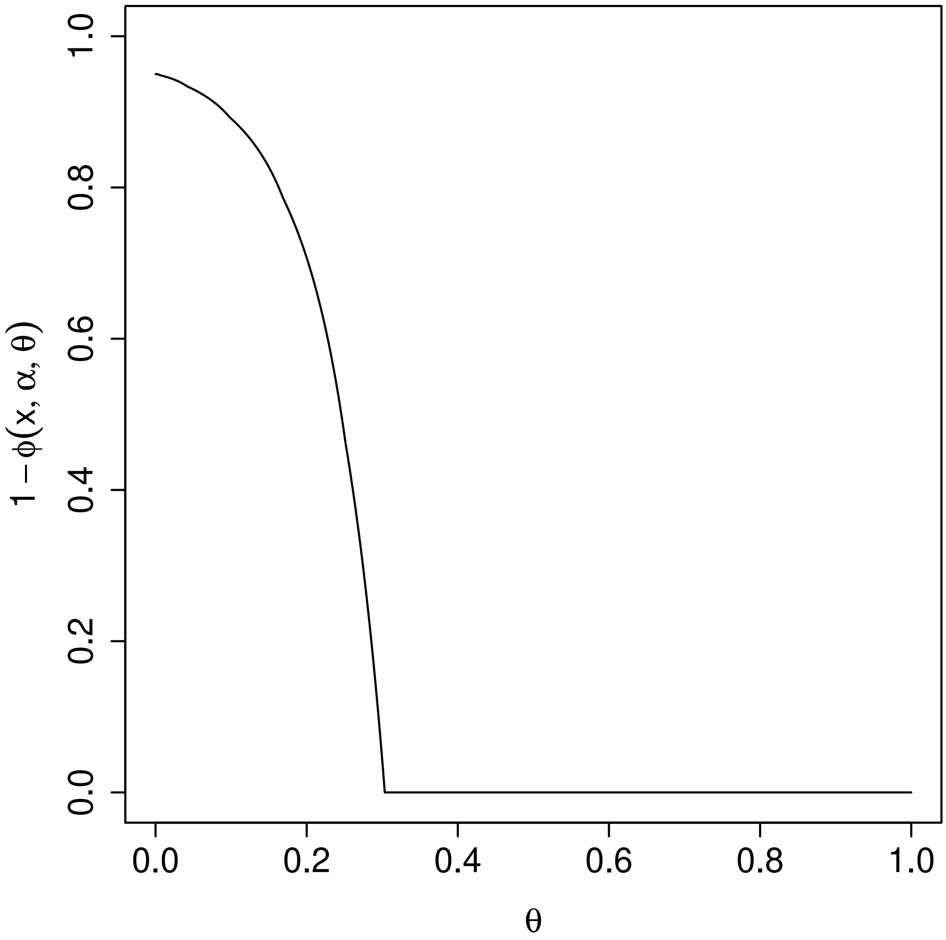
Data $x = 0$ (solid curve) $x = 4$ (dashed curve) and $x = 9$ (dotted curve).



Binomial Example

Sample size $n = 10$, fuzzy confidence interval associated with UMPU test, confidence level $1 - \alpha = 0.95$.

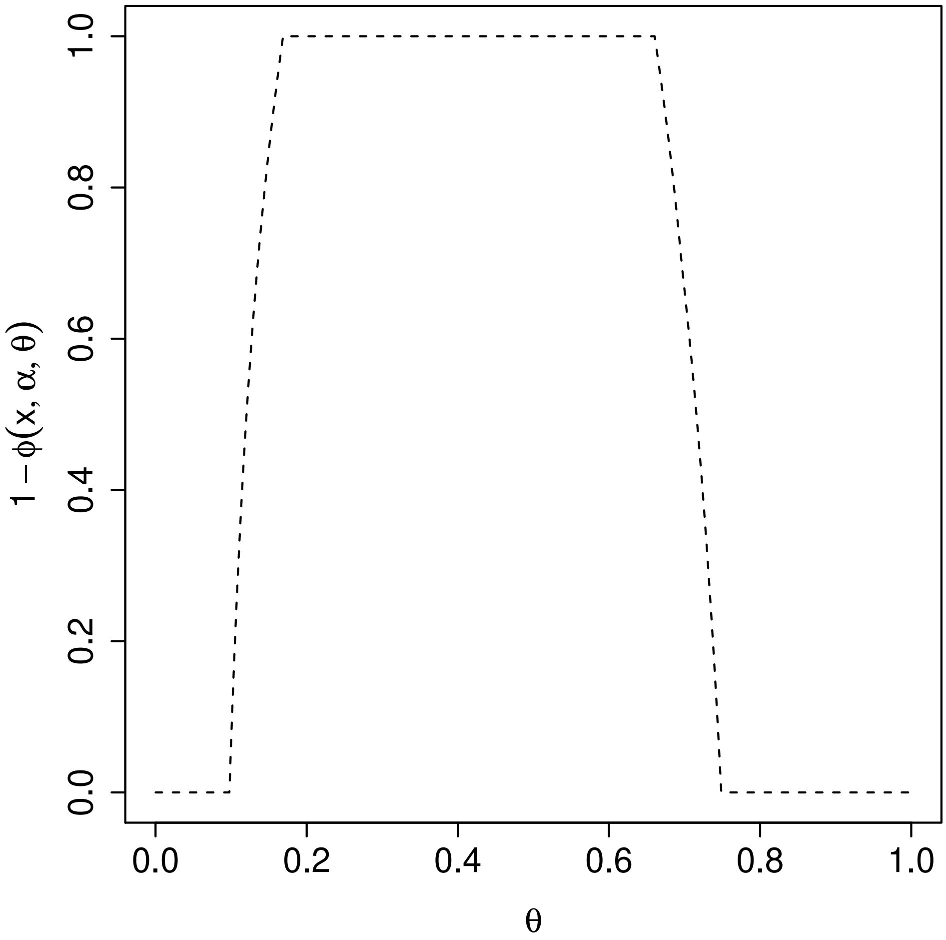
Data $x = 0$.



Binomial Example

Sample size $n = 10$, fuzzy confidence interval associated with UMPU test, confidence level $1 - \alpha = 0.95$.

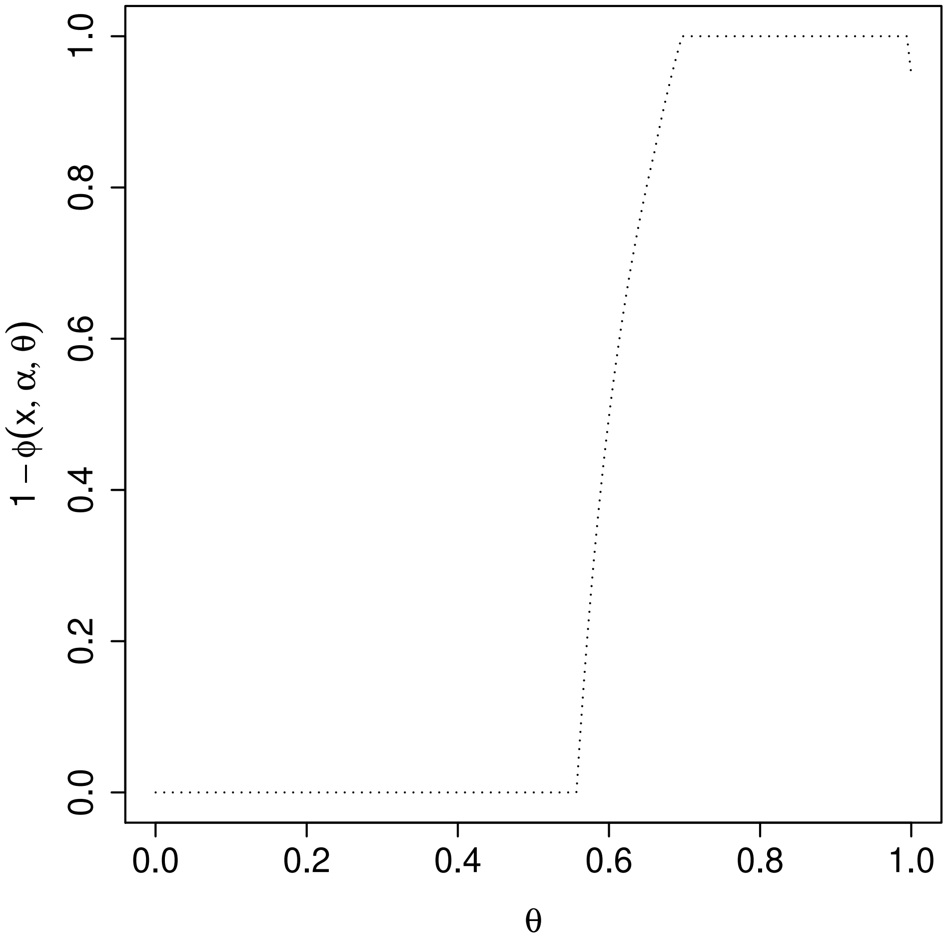
Data $x = 4$.



Binomial Example

Sample size $n = 10$, fuzzy confidence interval associated with UMPU test, confidence level $1 - \alpha = 0.95$.

Data $x = 9$.



Exactness

UMP and UMPU tests are *exact*

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

Fuzzy confidence intervals inherit exactness

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

Fuzzy P-values

For P-values to even be definable a test must have nested fuzzy critical regions,

$$\alpha_1 \le \alpha_2 \quad \text{implies} \quad \phi(x, \alpha_1, \theta) \le \phi(x, \alpha_2, \theta).$$

For any such test for discrete data and for any x and θ

$$\alpha \mapsto \phi(x, \alpha, \theta) \tag{$*$}$$

is a continuous non-decreasing function that maps $[0, 1]$ onto $[0, 1]$.

So $(*)$ is the cumulative distribution function of a continuous random variable P, which we call the *fuzzy P-value* of the test.

Exactness Again

UMP and UMPU tests are *exact*

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

By definition of fuzzy P-value

$$\text{pr}\{P \le \alpha \mid X\} = \phi(X, \alpha, \theta_0)$$

Hence fuzzy P-values also inherit exactness

$$\begin{aligned} \text{pr}_{\theta_0}\{P \le \alpha\} &= E_{\theta_0}\{\text{pr}\{P \le \alpha \mid X\}\} \\ &= E_{\theta_0}\{\phi(X, \alpha, \theta_0)\} \\ &= \alpha, \end{aligned}$$

Probability Density Function

Cumulative distribution of fuzzy P-value is

$$\alpha \mapsto \phi(x, \alpha, \theta_0) \qquad (*)$$

Probability density function is

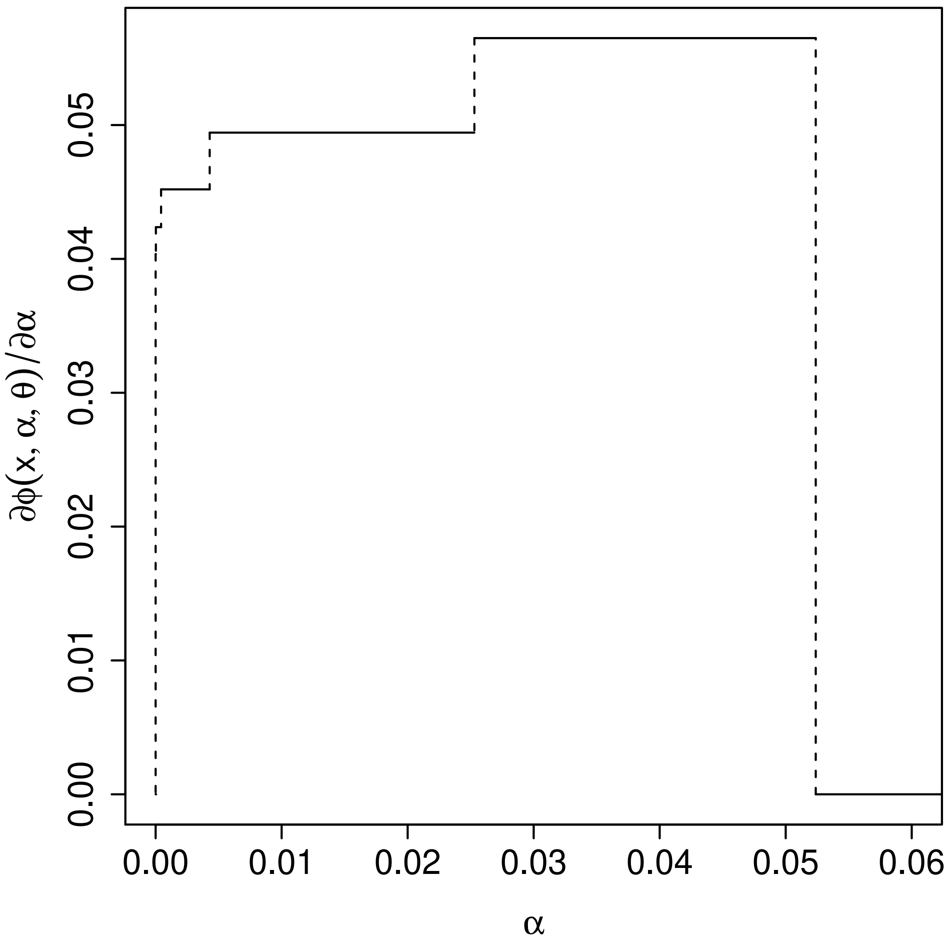
$$\alpha \mapsto \frac{\partial}{\partial \alpha} \phi(x, \alpha, \theta_0) \qquad (**)$$

For UMP tests, fuzzy P-values are uniformly distributed on an interval.

For UMPU tests, $(*)$ is piecewise linear and $(**)$ is piecewise constant (a step function).

Binomial Example

Sample size $n = 10$, fuzzy P-value associated with UMPU test, null hypothesis $\theta_0 = 0.7$, data $x = 10$.



Fuzzy and Randomized Concepts

Decisions

Fuzzy decision reports $\phi(x, \alpha, \theta_0)$.

Randomized decision generates Uniform$(0, 1)$ random variate U, and reports "reject H_0" if $U < \phi(x, \alpha, \theta_0)$ and "accept H_0" otherwise.

P-values

Fuzzy P-value is the theoretical random variable having the cumulative distribution function $\alpha \mapsto \phi(x, \alpha, \theta_0)$.

Randomized P-value is a realization (simulated value) of this random variable.

Fuzzy and Randomized Concepts (Continued)

γ-Cuts

If I_B is the membership function of a fuzzy set B, the *γ-cut* of B is the crisp set

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

Confidence Intervals

Fuzzy confidence interval is the fuzzy set B with membership function

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

Randomized confidence interval is the crisp set $^{U}I_B$, where U is uniform $(0, 1)$ random variate.

Situations with UMP and UMPU Tests

- Binomial
- Poisson
- Negative Binomial
- Two Binomials $p_1(1-p_2)/p_2(1-p_1)$
- Two Poissons μ_1/μ_2
- Two Negative Binomials $(1-p_2)/(1-p_1)$
- McNemar $p_{12}/(p_{12}+p_{21})$
- Fisher $p_{11}p_{22}/p_{12}p_{21}$

Summary

- Fuzzy decisions, confidence intervals, and P-values based on UMP and UMPU tests are the Right Thing (exact and uniformly most powerful).

- Crisp confidence intervals are the Wrong Thing (for discrete data).

- UMP and UMPU for any exponential family with single parameter of interest.

- Fuzzy outside classical UMP and UMPU?

- Fuzzy or Randomized?

Appendix: UMP

For one-parameter exponential family having canonical statistic $T(X)$ and canonical parameter θ there exists *UMP test* with hypotheses

$$H_0 = \{\, \vartheta : \vartheta \leq \theta \,\}$$
$$H_1 = \{\, \vartheta : \vartheta > \theta \,\}$$

significance level α, and critical function

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

where γ and C are determined by

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

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

Appendix: UMPU

For one-parameter exponential family having canonical statistic $T(X)$ and canonical parameter θ there exists *UMPU test* with hypotheses

$$H_0 = \{\, \vartheta : \vartheta = \theta \,\}$$
$$H_1 = \{\, \vartheta : \vartheta \neq \theta \,\}$$

significance level α, and critical function

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

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

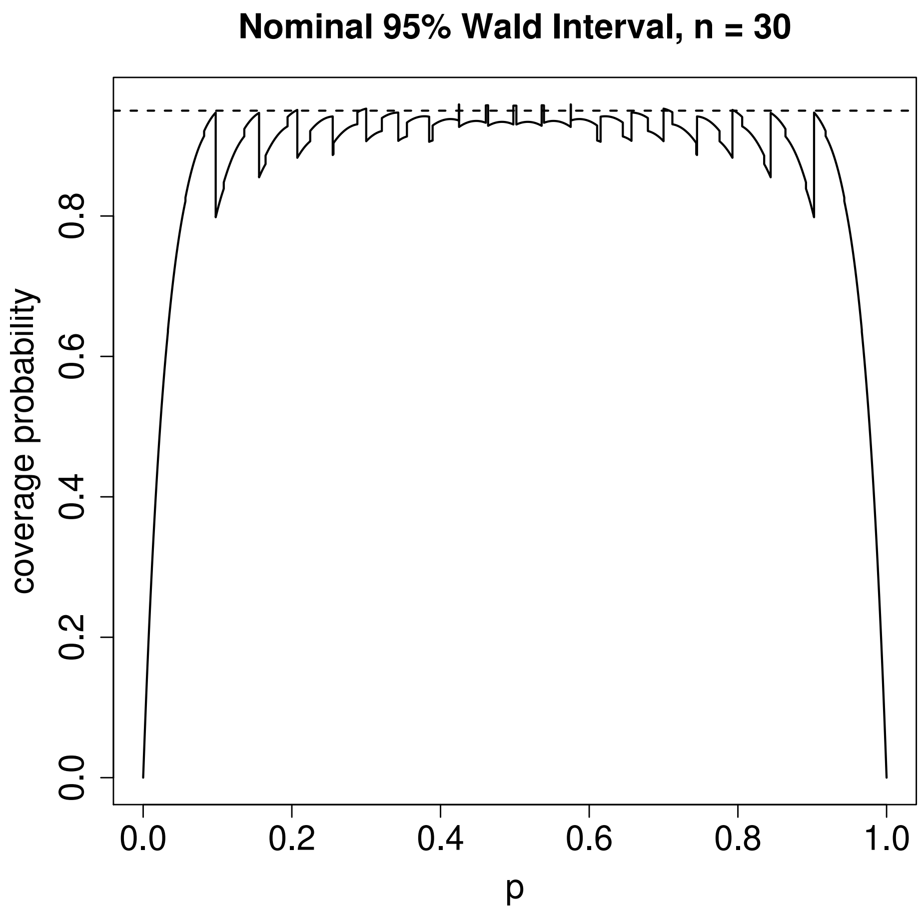
$$E_\theta\{\phi(X, \alpha, \theta)\} = \alpha$$
$$E_\theta\{T(X)\phi(X, \alpha, \theta)\} = \alpha E_\theta\{T(X)\}$$

Appendix: More Bad Crisp Intervals I

Performance of usual (Wald) interval

$$\hat{p} \pm 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

for Binomial$(30, p)$. Dotted line is nominal level (0.95).

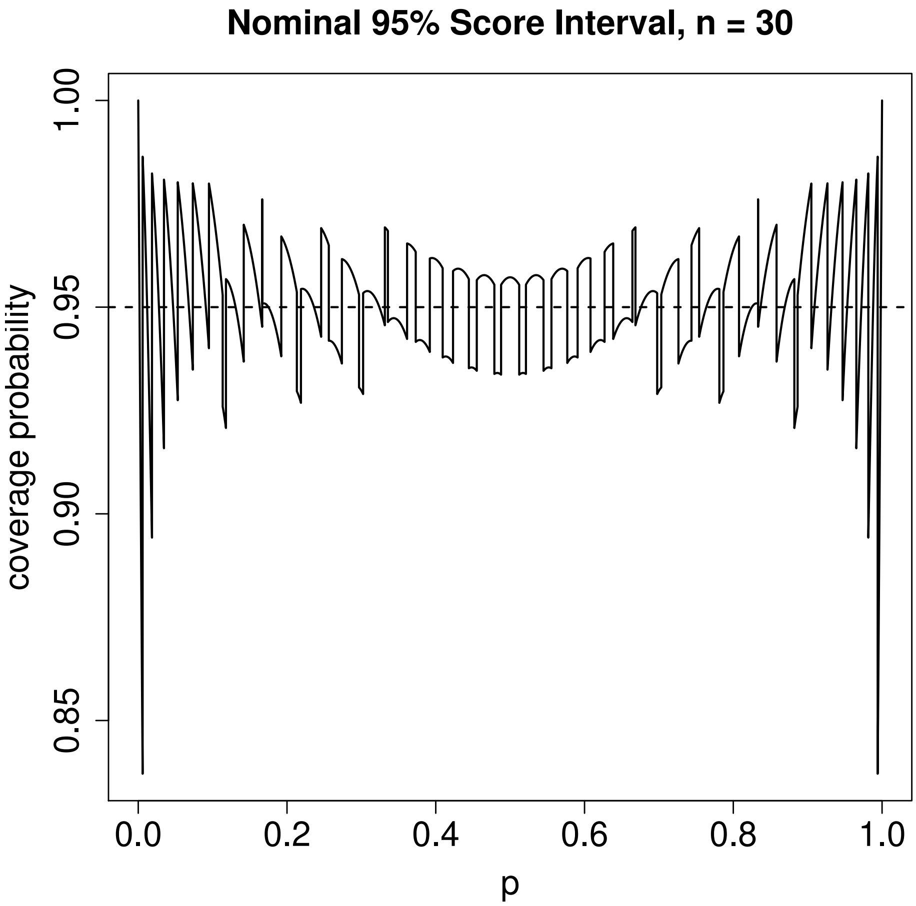


Appendix: More Bad Crisp Intervals II

Performance of score interval

$$\left\{ p : |\hat{p} - p| < 1.96 \sqrt{\frac{p(1-p)}{n}} \right\}$$

for Binomial$(30, p)$. Dotted line is nominal level (0.95).

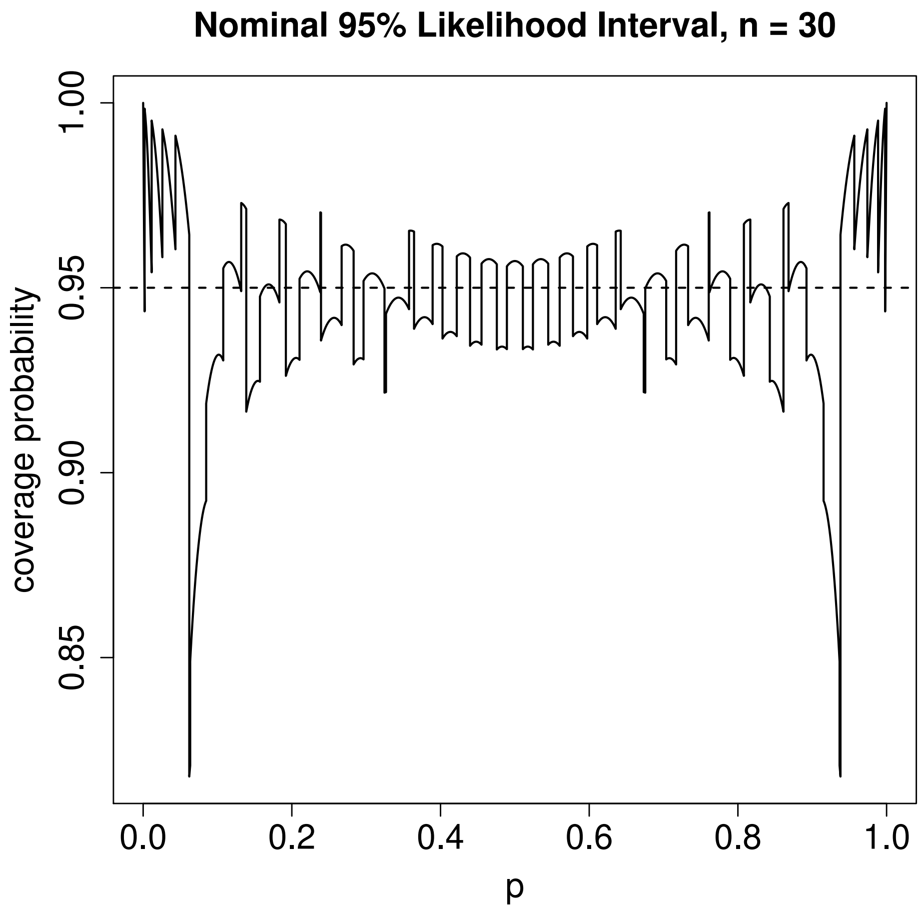


Appendix: More Bad Crisp Intervals III

Performance of likelihood interval

$$\left\{ p : 2\Big[l_n(\widehat{p}) - l_n(p) \Big] < 1.96^2 \right\}$$

where l_n is log likelihood for Binomial$(30, p)$. Dotted line is nominal level (0.95).

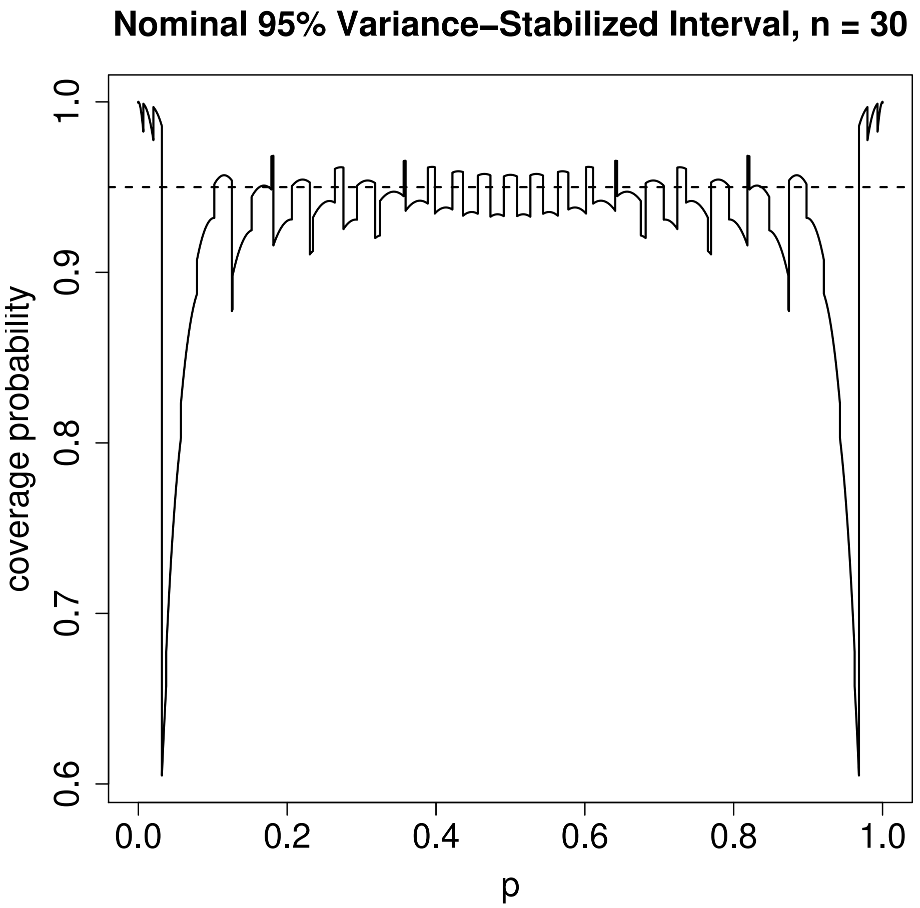


Appendix: More Bad Crisp Intervals IV

Performance of variance stabilized interval

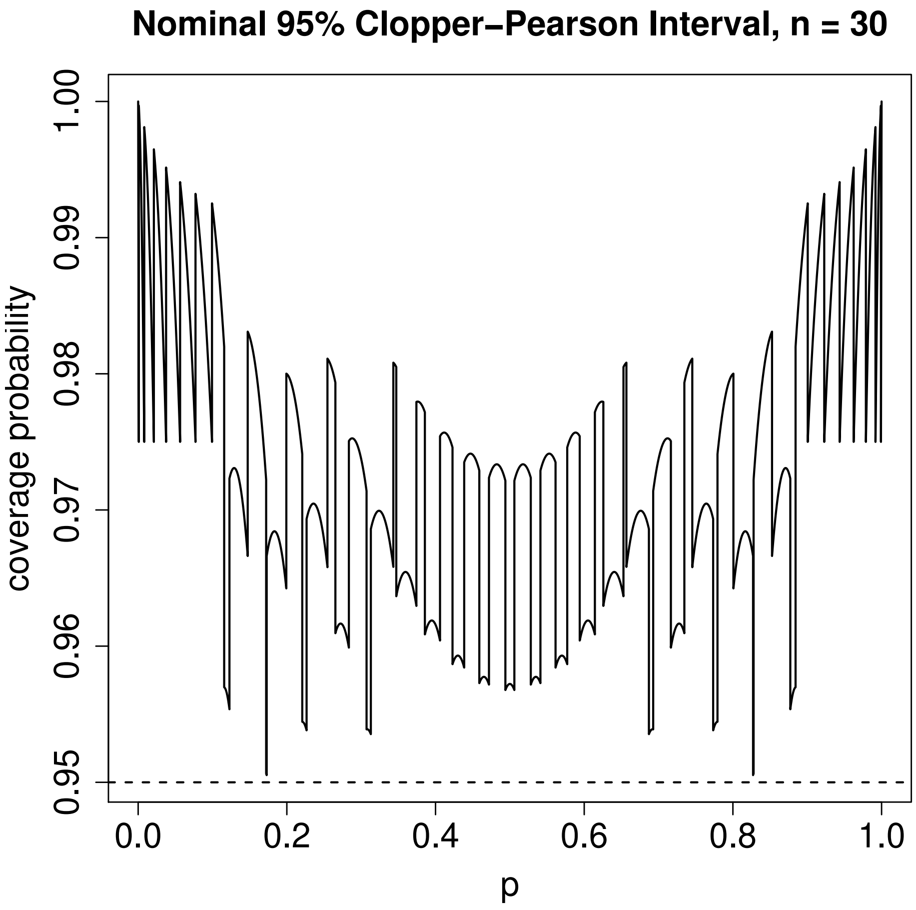
$$g^{-1}\left(g(\widehat{p}) \pm \frac{1.96}{\sqrt{n}}\right)$$

where $g(p) = 2\sin^{-1}(\sqrt{p})$ for Binomial$(30, p)$. Dotted line is nominal level (0.95).



Appendix: More Bad Crisp Intervals V

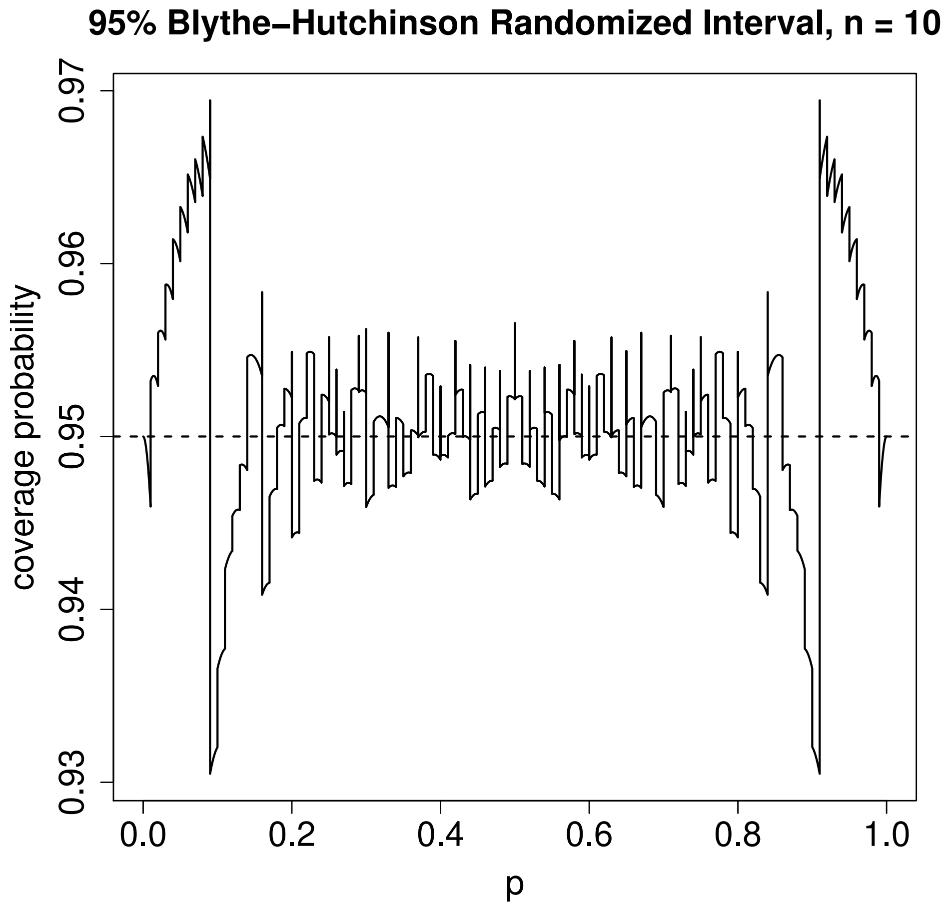
Performance of Clopper-Pearson "exact" interval for Binomial$(30, p)$. Dotted line is nominal level (0.95).



Appendix: An Early Randomized Interval

Performance of Blyth-Hutchinson randomized interval for Binomial$(30, p)$. Dotted line is nominal level (0.95).

Would be exact except for rounding error. Randomization variate U rounded to one sig. fig. Interval endpoints rounded to two sig. fig.



Appendix: Old Literature

Blyth and Hutchinson (*Biometrika*, 1960)
Table of Neyman-shortest unbiased confidence intervals for the binomial parameter.

Lehmann and Scheffé (*Sankyā*, 1950, 1955)
Completeness, similar regions, and unbiased estimation.

Eudey (*Berkeley Tech. Rept.*, 1949)
On the treatment of discontinuous random variables.

Wald (*Econometrica, 1947*)
Foundations of a General Theory of Sequential Decision Functions.

von Neumann and Morgenstern (1944)
Theory of Games and Economic Behavior.